Translated Poisson Approximation with Stein's Method

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Adrian Röllin

von

Freienbach SZ

Promotionskomitee

Prof. Dr. Andrew Barbour (Vorsitz)
Prof. Dr. Erwin Bolthausen
Prof. Dr. Louis H. Y. Chen (Singapur)

Zürich, 2006

Zusammenfassung

Charles Stein hat in seinem Artikel aus dem Jahre 1972 eine Methode vorgestellt, um Summen von abhängigen Zufallsvariablen mit der Normalverteilung in der Kolmogorov-Metrik zu approximieren und erhält Resultate ganz im Sinne der klassischen Resultate von Berry (1941) und Esseen (1942). Der von Stein benutzte Ansatz hat sich hauptsächlich aus drei Gründen als sehr fruchtbar erwiesen. Erstens kann die Methode an die verschiedensten Arten von Abhängigkeiten der Summanden angepasst werden; die Resultate von Bolthausen (1984) stellen hier sicher einen Höhepunkt in der Normalapproximation dar. Weiter können auch viele andere Verteilungen erfolgreich mit der Methode behandelt werden. Als erstes wurde die Methode von Chen (1975) für die Poissonverteilung angewandt; hier gibt Barbour et al. (1992) eine gute Übersicht über die vielfältigen Resultate. In den späten 80er Jahren und in den 90er Jahren folgte die Anpassung auf andere Verteilungen wie die multivariate Normalverteilung durch Barbour (1990) und Götze (1991), Gammaverteilung durch Luk (1994), Binomialverteilung durch Ehm (1991) und viele mehr. Und als dritten Vorteil sei die Möglichkeit erwähnt, Konvergenzraten in verschiedensten Metriken zu erhalten, wovon die Kolmogorov-Metrik, die Wasserstein-Metrik (auch Lipschitz-Metrik genannt) und die Totalvariationsmetrik nur einige wenige Beispiele sind.

In den drei hier vorgestellten Artikeln wird im speziellen auf zwei Verteilungen eingegangen: Die verschobene Poisson- und die zentrierte und symmetrische Binomialverteilung. Der Grund, diese Verteilungen zu betrachten, ist einfach einzusehen. Angenommen, man möchte eine Zufallsvariable W, welche nur Werte auf den ganzen Zahlen $\mathbb{Z}$ annimmt, durch eine einfachere Verteilung approximieren. Ist W eine Summe von Zufallsvariablen, so bietet sich natürlich im Rahmen eines zentralen Grenzwertsatzes die Normalverteilung an. Dies kann letztlich jedoch nicht befriedigend sein, da die Normalverteilung stetig ist, unser Problem jedoch diskret. Eine Approximation wird also nur mit Metriken sinnvoll sein, welche diesen Unterschied weitgehend unberücksichtigt lassen, wie z.B. die Kolmogorov-Metrik oder die Wasserstein-Metrik.

Aus diesen Überlegungen heraus versuchen wir einfache Verteilungen zu benutzen, welche als Ersatz für die Normalverteilung im Diskreten dienen und dieselbe glockenförmige Form aufweisen sollen. Es stellt sich heraus, dass die Poisson- und die Binomialverteilung im Rahmen der Steinschen Methode besonders einfach handzuhaben sind. Diese Verteilungen sollen uns als Grundbaustein für „quasi"-zentrale Grenzwertsätze im Diskreten dienen. Da wir den Mittelwert und die Varianz der genannten Verteilungen nicht beliebig steuern können, betrachten wir stattdessen Verschiebungen dieser Verteilungen um den Mittelwert, d.h. wir wählen die Parameter so, dass die Varianz gleich der Varianz unserer Variable W ist und verschieben dann die Verteilung so weit, dass auch der Mittelwert mit W übereinstimmt (d.h. insbesondere, dass wir nicht W skalieren wie im zentralen Grenzwertsatz üblich, sondern die Verteilung, mit der wir approximieren; in diesem Sinne ist das Wort „quasi" zu verstehen, da mit $\sigma^2 \to \infty$ natürlich keine wirkliche Konvergenz stattfindet). Weiter verwenden wir statt der Metriken, die wir oben erwähnt haben, die Totalvariationsmetrik, welche im Diskreten die natürlichste Wahl ist. Wir betrach-

ten auch eine lokale Metrik, mit welcher lokale Grenzwertsätze abgeleitet werden können.

In den Beweisen von zentralen Grenzwertsätzen in der Kolmogorov-Metrik fliesst üblicherweise eine sogenannte Glattheitseigenschaft ein. Diese kann z.B. in der Form einer Konzentrationsungleichung durch eine Abschätzung der Wahrscheinlichkeiten

$$\mathbb{P}[a \leqslant W \leqslant b] \tag{1}$$

gegeben sein; siehe dazu z.B. Chen and Shao (2005). Eine Abschätzung des obigen Ausdrucks für alle $a, b \in \mathbb{R}$ mit $a - b < c$ kann auch aufgefasst werden als Abschätzung von

$$d_{\mathrm{K}}\bigl(\mathscr{L}(W), \mathscr{L}(W + c)\bigr),$$

wobei d_{K} die Kolmogorov-Metrik und $\mathscr{L}(W)$ die Verteilung von W sei. Es stellt sich heraus, dass, wenn wir im Rahmen der Steinschen Methode mit diskreten Verteilungen und der Totalvariationsmetrik arbeiten, die analoge Grösse

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathscr{L}(W + 1)\bigr), \tag{2}$$

abgeschätzt werden muss, wobei d_{TV} die Totalvariationsmetrik sei. Solche Grössen werden in der Regeln mit Hilfe von Coupling-Methoden geschätzt, was jedoch in der Regel nur dann gelingt, wenn W eine Summe von unabhängigen Zufallsvariablen ist.

Die führt uns zum Ansatz des ersten Artikels, der bereits von MacDonald (1979) angewandt wurde. Angenommen, man habe eine zusätzliche Zufallsgrösse X auf demselben Wahrscheinlichkeitsraum definiert wie W (X könnte z.B. einfach ein Teil der Summanden von W sein). Wenn sich nun die Verteilung von W bedingt auf X als Summe von unabhängigen Summanden darstellen lässt, so erhalten wir eine Abschätzung von (2), welche nun einfach das Supremum aller Abschätzungen von (2) bezüglich der bedingten Verteilung von W und aller möglichen Werte von X ist. Was im Wesentlichen übrig bleibt ist zu zeigen, dass der bedingte Erwartungswert von W bezüglich X in etwa normalverteilt ist. D.h. wir stellen W dar als etwas ungefähr normalverteiltes plus eine Summe von unabhängigen Zufallsvariablen, welche erstens selber natürlich auch in etwa normalverteilt sind, zusätzlich aber noch garantieren, dass (2) klein ist. Dies ermöglicht dann in der Tat eine Approximation mit einer verschobenen Poissonverteilung. Die Raten, welche in den Beispielen erreicht werden, entsprechen denen des klassischen Theorems von Berry und Esseen für Summen von unabhängigen Zufallsvariablen bezüglich der Kolmogorov-Metrik.

Der zweite Artikel verfolgt einen etwas spezifischeren Ansatz, in welchem W die Summe von lokal abhängigen Summanden ist. Dieser Fall ist zwar durch den ersten Artikel abgedeckt, erlaubt aber direktere Berechnungen. Zudem wird die verschobene Poissonverteilung durch eine zentrierte und symmetrische Binomialverteilung ersetzt, was gewisse technische Vorteile hat und auch bessere Konstanten liefert, aber bezüglich der uns interessierenden Resultate gleichwertig zur verschobenen Poissonverteilung ist. Ähnlich wie in der Normalapproximation, macht man eine Taylorentwicklung um die einzelnen Summanden herum. Eine Abschätzung von (2) wird nun nicht mehr bezüglich ganz W benötigt, sondern jeweils von W bedingt auf den Summanden, um den man entwickelt hat. Ähnlich wie im ersten Artikel lässt sich (2) nun abschätzen, wenn wir zeigen, dass sich W, bedingt auf den Summanden und eventuell auf eine zusätzliche Zufallsvariable, darstellen lässt als Summe von unabhängigen Zufallsvariablen.

Der dritte Artikel verfolgt einen ganz anderen Ansatz, welcher von Stein (1986) für die Normalverteilung eingeführt wurde. Dabei konstruiert man zu einer gegebenen Zufallsvariable W eine weitere Zufallsvariable W', so dass (W, W') und (W', W) dieselbe Verteilung haben und so dass

$$\mathbb{E}^W W' = (1 - \lambda)W$$

für ein $\lambda > 0$ gilt. Man kann nun ein Theorem für eine Normalapproximation formulieren welches effiziente Resultate liefert, wenn $\mathbb{E}^W (W' - W)^2$ nicht zu stark fluktuiert, d.h. wenn die zu erwartende Abweichung $|W' - W|$ von W für alle möglichen Werte von W in etwa gleich gross ist. Führt man nun zusätzlich die neue Bedingung ein, dass fast sicher

$$W' - W \in \{-1, 0, 1\}, \tag{3}$$

so erhält man in der Tat Resultate in der Totalvariationsmetrik für die verschobene Poissonverteilung. Obschon in den Berechnungen nirgends eine explizite Abschätzung von (2) vorkommt, so wird doch klar, dass die Bedingung (3) implizit den Effekt hat, dass (2) klein muss.

LITERATUR

A. D. Barbour (1990). Stein's method for diffusion approximations. *Probab. Th. Rel. Fields* **84**, 297–322.

A.D. Barbour, L. Holst, and S. Janson (1992). *Poisson Approximation*. Oxford University Press.

A. C. Berry (1941). The accuracy of the Gaussian approximation to the sum of indeendent variates. *Trans. Amer. Math. Soc.* **49**, 122–136.

E. Bolthausen (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**, 379–386.

L.H.Y. Chen (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–45.

L.H.Y. Chen and Q.-M. Shao (2005). Stein's method for normal approximation. In *An Introduction to Stein's method*, Lecture Notes Series. Institute for Mathematical Sciences, National University of Singapore.

W. Ehm (1991). Binomial approximation to the Possion Binomial distribution. *Statist. Probab. Lett.* **11**, 7–16.

C-G. Esseen (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat., Astro. Fysik.* **28A**, 1–19.

F. Götze (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.

H. M. Luk (1994). *Stein's method for the Gamma distribution and related statistical applications.* PhD dissertation, USC.

D. McDonald (1979). On local limit theorem for integer valued random variables. *Theory Probab. Appl.* **24**, 613–619.

Y. Rinott and V. Rotar (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7**, 1080–1105.

C. Stein (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 583–602.

C. Stein (1986). *Approximate computation of expectations*, IMS Lecture Notes. Institute of Mathematical Statistics, Hayward.

Summary

In the year 1972, Charles Stein presented a new method for approximating sums of dependent random variables by a normal distribution with respect to the Kolmogorov metric, and obtained results much in the spirit of the classical results of Berry (1941) and Esseen (1942). Stein's approach has proven fruitful mainly for three reasons. First, the method can be adapted to a variety of settings; the results of Bolthausen (1984) represent a highlight in normal approximation. Second, many other distributions can be successfully treated by the method. It was first adapted to the Poisson distribution by Chen (1975); see Barbour et al. (1992) for a good survey of the various results. In the late 80s and in the 90s the method was adapted to many other distributions, such as the multivariate normal distribution by Barbour (1990) and Götze (1991), the gamma distribution by Luk (1994), the binomial distribution by Ehm (1991) and many more. And third, estimates of rates of convergence may be obtained in a variety of metrics, where the Kolmogorov metric, the Wasserstein (Lipschitz) metric, total variation metric are only a few examples.

In the present three papers, we concentrate on two distributions: the translated Poisson and centred and symmetric binomial distributions. The reason for considering these distributions is the following. Assume that one wants to approximate a random variable W, taking only values on the integers $\mathbb{Z}$, by a simpler distribution. If W is a sum of random variables, we first try of course with the normal distribution in the sense of a central limit theorem. However, this is not fully satisfying, as the normal distribution is continuous, but W discrete. Thus, we can obtain meaningful approximations only if we use metrics which are insensitive to this difference, such as the Kolmogorov or the Wasserstein metric.

It is now but a short step to try to find distributions that serve as substitute for the normal distribution in the discrete setting and that exhibit a similar bell shaped form. It turns out that the Poisson and the binomial distributions are particular easy to handle with Stein's method. These distributions serve us as basis for obtaining 'quasi' central limit theorems in the discrete setting. As we cannot control both the mean and the variance of these distributions independently, we consider translations instead; that is, we choose the parameters so that the variance is equal to the variance of W and then translate the distribution to fit the mean of W as well (note that we do not scale W as is usual done in the central limit theorem, but the distribution by which we approximate; this is how we should understand 'quasi', as with $\sigma^2 \to \infty$ there is of course no limiting distribution and thus no convergence taking place). Further, instead of the above mentioned metrics we use the total variation metric, which is the most natural choice in the discrete setting. We also consider a local metric, from which local limit theorems may be obtained.

In the proofs of central limit theorems in the Kolmogorov metric, there is usually a smoothing inequality involved. Such an inequality can, for instance, take the form of a concentration inequality; that is, an estimate of the probabilities

$$\mathbb{P}[a \leqslant W \leqslant b], \tag{1}$$

as in Chen and Shao (2005). A bound on the above term for all $a, b \in \mathbb{R}$ with $a - b < c$ can also be seen as a bound on

$$d_{\mathrm{K}}\big(\mathscr{L}(W), \mathscr{L}(W + c)\big),$$

where d_{K} denotes the Kolmogorov metric and $\mathscr{L}(W)$ is the distribution of W. It turns out that, if we work within the framework of Stein's method in the discrete setting, the analogous expression

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathscr{L}(W + 1)\big) \tag{2}$$

has to be estimated, where d_{TV} denotes the total variation metric. Such expressions are usually bounded using coupling methods, which, however, can in general be applied only if W is a sum of independent random variables.

This leads us to the approach of the first paper, already used by MacDonald (1979). Assume that we are given a random element X on the same probability space as W (for instance, X could be some of the summands of W). Now, if the distribution of W given X can be represented as a sum of independent random variables, we obtain an estimate of (2), which is just the supremum of all the estimates of (2) with respect to the conditional distributions of W given all the possible values of X. What essentially remains is to show that the conditional expectation of W given X, is approximately normally distributed. That is, we have reduced the distribution of W to something which is approximately normal plus a sum of independent random variables, which themselves are of course also approximately normal, but guarantee in addition that (2) is small. This allows us indeed to deduce an approximation theorem. The rates achieved in the examples correspond to those obtained by the classical Berry-Esseen theorem for sums of independent random variables with respect to the Kolmogorov metric, but now in total variation and in a dependent setting.

The second paper considers a somewhat more specific approach, in which W is a sum of locally dependent random variables. Although this case is in fact covered by the first paper, it allows more direct calculation. Moreover, we replace the translated Poisson distribution by a centred and symmetric binomial distribution, which has some technical advantages and yields better constants, but which is equivalent to the former as far as what is of interest to us. As for the normal approximation, we conduct a Taylor expansion about the individual summands. A good bound on (2) is needed now not of the whole of W, but of W conditioned on the summands about which we have expanded. As in the first paper, one can obtain bounds on (2) if it can be shown that W, conditioned on a summand and, if necessary, also on an additional random variable, can be represented as a sum of independent summands.

The third paper pursues a very different line of argument, one originally introduced by Stein (1986) for the normal distribution. There, one constructs for given W another random variable W' in such a way that (W, W') and (W', W) are equal in distribution and such that

$$\mathbb{E}^W W' = (1 - \lambda)W$$

for some $\lambda > 0$. With this, it is possible to formulate an approximation theorem in which convergence is obtained if $\mathbb{E}^W(W' - W)^2$ does not fluctuate to much; that is, if the expected deviation $|W' - W|$ from W is about the same for all possible values of W. If one introduces now the additional condition that almost surely

$$W' - W \in \{-1, 0, 1\}, \tag{3}$$

we obtain approximation results for the translated Poisson distribution in total variation. Although quantities of the form (2) are not directly involved in the calculations, it becomes clear that condition (3) implicitly has the effect that (2) must be small.

Danksagung

Viele Leute haben mich auf dem Weg meiner Dissertation begleitet.

Ich bin zuallererst meinem Doktorvater Professor Andrew Barbour zu tiefem Dank verpflichtet. Seine Ideen und Anregungen haben meine Arbeit wesentlich beeinflusst und mitbestimmt und durch seine menschliche und humorvolle Art hat er über all die Jahre eine herrvorragende Arbeitsatmosphäre geschaffen. Ich danke ihm ganz herzlich für die vielen Verbesserungsvorschläge in meinen Artikeln und seine geduldigen Korrekturen meines Englischs.

Ganz herzlich danken möchte ich Professor Erwin Bolthausen und Professor Louis Chen für die Begutachtung der Dissertation.

Danken möchte ich meinen Kolleginnen und Kollegen am Institut. Die abwechslungsreichen und teilweise hitzigen Diskussionen über mathematische und alltägliche Sichten und Ansichten haben das Leben in den Arbeitspausen nicht immer erholsam, sicherlich aber oft spannend und unterhaltsam gemacht. Besonders erwähnen möchte ich zwei Kollegen aus unserer Forschungsgruppe: Bruno Nietlispach, mit dem ich über vier Jahre lang das Büro geteilt habe, und Dominic Schuhmacher. Mit ihnen habe ich viele Diskussionen geführt, und wir haben uns gegenseitig gelehrt und belehrt.

Ich danke auch ganz herzlich den vielen Personen am Institut, welche mir bei administrativen und technischen Fragen immer hilfreich zur Seite gestanden haben.

Am allermeisten danken möchte ich meinen Eltern und meiner Freundin Alexandra, die mich all die Jahre bedingungslos unterstützt haben und ohne die ich nicht hier angelangt wäre, wo ich jetzt bin.

Adrian Röllin
Zürich, Oktober 2006

Contents

INTRODUCTION

1. Limits to the normal distribution

In the sequel, I give a short (and therefore incomplete) overview of the development of the central and local limit theorems. For the central limit theorem there are many books giving historical surveys and detailed insight into the different methods of proofs. I want to mention explicitly Petrov (1975), Prokhorov and Statulevičius (2000) and Adams (1974).

1.1. The central limit theorem. Let $\xi = (\xi_1, \xi_2, \dots)$ be a sequence of random variables and write $S_n = \sum_{i=1}^n \xi_n$. We say that ξ satisfies an (integral) *central limit theorem* (CLT) if, with $\mu_n := \mathbb{E}S_n$ and $\sigma_n^2 := \operatorname{Var} S_n$,

$$\frac{S_n - \mu_n}{\sigma_n} \quad \xrightarrow{\mathscr{D}} \quad \mathcal{N}(0,1)$$

as $n \to \infty$, where $\mathcal{N}(0,1)$ is the standard normal distribution. We will from now on assume that $\sigma_n^2 \to \infty$ as $n \to \infty$. For further use, define $\hat{\xi}_i = (\xi_i - \mathbb{E}\xi_i)/\sigma_n$ and $\hat{S}_n := \sum_{i=1}^n \hat{\xi}_i$. The investigation of conditions on ξ under which a CLT is satisfied has a long tradition in probability theory, and is of great importance for hypothesis testing in statistics. Probably the first results were obtained by Abraham de Moivre and Pierre Simon de Laplace in the 18th century, for the case in which the ξ_i are i.i.d. Bernoulli experiments. The first rigorous treatment of the CLT was initiated by Pafnuti Chebishev and his students Andrei Markov and Alexander Liapunov. Liapunov (1900, 1901) first introduced the method of characteristic functions.

This new approach resulted in the celebrated theorems by Berry (1941) and Esseen (1942) for sums of independent ξ_i. They were not only able to prove convergence itself, but were also the first to explicitly bound the accuracy of approximation for n fixed, with the help of inequality (2) below.

Since then, much has been done. In the case of real valued ξ_i, three important methods of proving CLTs have turned out to be fruitful because they yield not only convergence, but can also be used to obtain rates of convergence in various probability metrics. The most important one has been—and probably still is—the method of characteristic functions. Define the characteristic function of a random variable X as

$$f_X(t) := \mathbb{E}\mathrm{e}^{itX}.$$

Now, noting that $f_Z(t) = \mathrm{e}^{-t^2/2}$ for $Z \sim \mathcal{N}(0,1)$, one can then use the key fact that a CLT holds if (and only if)

$$f_{\hat{S}_n}(t) \to \mathrm{e}^{-t^2/2} \tag{1}$$

pointwise for all t. If, in addition, a rate of convergence is desired, one uses the *Esseen inequality* (also called the *smoothing lemma*), which states that

$$\bigl|\mathbb{P}[\hat{S}_n < x] - \Phi(x)\bigr| \leqslant C_1 \int_{-T}^{T} \frac{|f_{\hat{S}_n}(t) - \mathrm{e}^{-t^2/2}|}{|t|} dt + \frac{C_2}{T} \tag{2}$$

for any $T > 0$ and absolute constants C_1 and C_2, and where Φ is the distribution function of $\mathcal{N}(0,1)$. Estimate (2) results from the Fourier inversion formula. There is a huge literature based on this approach, and there are results with and without rates both in independent and dependent settings. Among these I would emphasise the important contributions of Bernstein (1927), Hoeffding and Robbins (1948), Ibragimov (1967), Tihomirov (1980) and Heinrich (1982), because these papers consider in particular sums of dependent random variables, which will also be the setting in the papers of this thesis. Clearly, most of the results using characteristic functions are expressed in terms of the Kolmogorov metric (also called uniform metric)

$$d_{\mathrm{K}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \sup_{x\in\mathbb{R}} \bigl|\mathbb{P}[X \leqslant x] - \mathbb{P}[Y \leqslant x]\bigr|,$$

because of (2).

A second approach I would like to mention is *Lindeberg's method* which appeared first in Lindeberg (1922, 1920), see also Trotter (1959). Here, one tries to obtain direct estimates of expressions of the form

$$\mathbb{E}h(\hat{S}_n) - \mathbb{E}h(Z) \tag{3}$$

for a suitable set $\mathcal{F}$ of test functions $h \in \mathcal{F}$. If the set $\mathcal{F}$ is large enough, one can obtain convergence and also rates of convergence in the (semi-)metric defined by $\mathcal{F}$. The main idea is to split up Z into a sum $Z = \sum_{i=1}^{n} Z_i$, where the Z_i are independent and normally distributed with $\mathbb{E}Z_i = 0$ and $\operatorname{Var} Z_i = \operatorname{Var} \hat{\xi}_i$. Then, one writes the difference in (3) as a telescoping sum

$$h(\hat{S}_n) - h(Z) = \sum_{i=1}^{n} \Bigl[h(W_{n,i} + \hat{\xi}_i) - h(W_{n,i} + Z_i)\Bigr],$$

where $W_{n,i} = \sum_{k=1}^{i-1} \hat{\xi}_i + \sum_{k=i+1}^{n} Z_i$. Using a Taylor expansion up to the third derivative of h, one easily obtains convergence and even rates of convergence, but only in a metric which does not allow for a simple probabilistic interpretation. Thus, if one wants to estimate (3) for larger sets of test functions (say for the Kolmogorov metric), one will need estimates of probabilities of the form

$$\mathbb{P}[a \leqslant W \leqslant b] \leqslant C(a - b), \tag{4}$$

called *concentration inequalities*; these were first introduced in Bergström (1944), see also Sazonov (1981). Adaptions to dependent settings can for example be found in Butzer and Schulz (1985). It has also been successfully used to deduce CLTs in Banach spaces, see for example Osipov and Rotar (1984) and Račkauskas (1991).

The third method, which I will use in the papers of this thesis, was introduced by Stein (1972). As in the Lindeberg method, one wants to find estimates of (3). However, one does this in an indirect way through the equality

$$f'(x) - xf(x) = h(x) - \mathbb{E}h(Z) \tag{5}$$

where $f = f_h$ is chosen such that the equality holds for a given h. Taking expectation with respect to $\hat{S}_n$ one regains the difference (3), that is,

$$\mathbb{E}\big\{f'(\hat{S}_n) - \hat{S}_n f(\hat{S}_n)\big\} = \mathbb{E}h(\hat{S}_n) - \mathbb{E}h(Z) \tag{6}$$

The success of this approach, however, depends on whether the l.h.s of (6) is easier to estimate than the r.h.s., and for this, one will need that the solution f has 'nice' properties (depending in turn on the properties of h). It can be shown in fact that the solution of (5) satisfies

$$\begin{aligned} \|f\|_\infty &\leqslant C_1 \min\big\{\|h\|, \|h'\|\big\}, \\ \|f'\|_\infty &\leqslant C_2 \min\big\{\|h\|, \|h'\|\big\}, \\ \|f''\|_\infty &\leqslant C_3 \|h'\|, \end{aligned} \tag{7}$$

for absolute constants C_1, C_2 and C_3, where the last bound is of course only useful if h is smooth enough. That the solution f of equation (5) indeed satisfies these bounds is far from obvious; and if the expectation $\mathbb{E}h(Z)$ were not subtracted on the r.h.s., where $Z \sim \mathcal{N}(0,1)$, it would not. Now, using Taylor expansions, one obtains bounds on the l.h.s. of (6) in terms of the bounds of f and its derivatives, which in turn yields bounds in terms of h and its derivatives through (7). To obtain rates of convergence in the Kolmogorov metric, one will often need in addition estimates of the form (4); see for example Chen and Shao (2004). However, Bolthausen (1984) also obtains results in the Kolmogorov metric, but without the use of such concentration inequalities.

There are also other methods for proving central limit theorem, such as the method of moments, which has become important for parameter estimation in statistics, and the method of cumulants, introduced in Statuljavičjus (1961), which is strongly connected to the method of characteristic functions and which seems particularly useful for large deviation results (see e.g. Saulis and Statulevičius (1991)).

1.2. The local limit theorem. With the notation of the previous section, assume in addition that S_n takes values on the integers, and that lattice span cannot be chosen to be larger than 1 (otherwise, divide S_n by a suitable factor). Then we say that ξ satisfies a *local limit theorem* (LLT, often also called *local central limit theorem*) if

$$\lim_{n\to\infty} \big|\sigma_n \mathbb{P}[S_n = k] - \varphi\big((k-\mu_n)/\sigma_n\big)\big| = 0 \tag{8}$$

uniformly for all $k \in \mathbb{Z}$, where φ is the density of the standard normal distribution. The LLT has not gained nearly as much attention as the CLT. This may be due to the fact that LLTs are in general much harder to obtain, because often stronger conditions on the summands must be assumed.

Due to a method introduced by Khinchin (1949), the LLT has become a useful tool in statistical mechanics and the theory of Gibbs random fields; see also Khinchin (1960) and Dobrushin and Tirozzi (1977).

First results are again due to De Moivre and Laplace for the binomial distribution. Later, Gnedenko (1948) used the method of characteristic functions to obtain results for the i.i.d. case, which were generalised by Rozanov (1957); Kolmogorov (1949) proved an LLT for Markov chains. Their approach, using characteristic functions and the Fourier inversion theorem, is still the one most often used. With the equalities

$$\sigma_n \mathbb{P}[S_n = k] = \frac{1}{2\pi} \int_{-\pi\sigma_n}^{\pi\sigma_n} f_{\hat{S}_n}(t) \mathrm{e}^{-it\frac{k-\mu_n}{\sigma_n}} dt,$$

$$\varphi\big((k-\mu)/\sigma_n\big) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{t^2}{2} - it\frac{k-\mu_n}{\sigma_n}} dt,$$

one can directly estimate the difference in (8) in terms of the characteristic functions.

There is also important work by Statuljavičjus (1961), who was already mentioned in the previous section.

One can also obtain LLTs using Stein's method. This approach is one of the main topics of this thesis. I will discuss this in the next section.

2. Discrete Gaussian-like distributions

As has become clear from the previous discussion of the LLT, the case of integer valued summands in a CLT is special. The approximation of the probabilities of S_n with the density of the normal distribution in an LLT yields more information about the distribution of S_n, but it remains unsatisfactory, due to the different supports of the distributions involved. It is therefore natural to look for integer valued distributions, which could serve as an alternative to the normal distribution. Of course, one would have to require several properties of such a distribution: the bell-shape of the normal distribution, with at most exponential tails, and the possibility to adjust the mean and the variance independently. The latter is important, because if we want to approximate S_n, we will not scale it, but rather adjust the parameters of our distribution such that it has the same variance as S_n, much as in the LLT.

As a first attempt, one could try the distribution of $\lfloor Z_n \rfloor$, where $Z_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$. It seems however difficult to work with this distribution within the framework of Stein's method. A more natural choice is to take a Poisson (or a symmetric Binomial) distribution, with the parameters set so that the variance is as desired, and then to shift this distribution accordingly to fit the mean. A rounding problem will of course appear, as the shift has to be integer-valued, but this turns out to be just a minor technicality.

2.1. The metrics. Although, one can still obtain results in the Kolmogorov metric or the Wasserstein-metric[1], the total variation metric d_{TV} is a more natural

[1]Recall that the Wasserstein-metric can be defined as
$$d_{\mathrm{W}}\big(\mathscr{L}(X), \mathscr{L}(Y)\big) = \textstyle\int_{-\infty}^{\infty} \big|\mathbb{P}[X \leqslant x] - \mathbb{P}[Y \leqslant x]\big| dx$$

choice for the comparison of distributions on the integers, as it is simply described in terms of the point probabilities. For given integer-valued random variables X and Y with point probabilities $\{p_k, k \in \mathbb{Z}\}$ and $\{q_k, k \in \mathbb{Z}\}$, respectively, we define

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |p_k - q_k|.$$

This metric also has an important coupling interpretation; we have

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \sup_{(X', Y')} \mathbb{P}[X' \neq Y'],$$

where the supremum is taken over all possible couplings of $\mathscr{L}(X)$ and $\mathscr{L}(Y)$. A third way of describing this metric is in terms of test functions. Define $\mathcal{F}_{\mathrm{TV}}$ to be the set of all indicator functions on $\mathbb{Z}$. Then one can show that

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \sup_{h \in \mathcal{F}_{\mathrm{TV}}} \bigl|\mathbb{E}h(X) - \mathbb{E}h(Y)\bigr|,$$

which allows now to apply methods that are based on the estimation of (3).

As for the LLT, we can also ask for the local accuracy of approximation. To this end we define the metric

$$d_{\mathrm{loc}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \sup_{k \in \mathbb{Z}} |p_k - q_k|,$$

which can also be represented in the form

$$d_{\mathrm{loc}}\bigl(\mathscr{L}(X), \mathscr{L}(Y)\bigr) = \sup_{h \in \mathcal{F}_{\mathrm{loc}}} \bigl|\mathbb{E}h(X) - \mathbb{E}h(Y)\bigr|,$$

where $\mathcal{F}_{\mathrm{loc}}$ is the set of all indicators on $\mathbb{Z}$, which attain the value 1 in exactly one point and are 0 otherwise.

If now Y_n has a translated Poisson or a centred and shifted binomial distribution with the mean and variance of S_n, it is clear that an LLT holds for Y_n, as both distributions can be respresented as a sum of i.i.d. random variables, which allows the application of classical results, for example Gnedenko (1948). To prove an LLT for S_n, it is therefore enough to show that

$$d_{\mathrm{loc}}\bigl(\mathscr{L}(S_n), \mathscr{L}(Y_n)\bigr) = \mathrm{o}(\sigma_n^{-1}). \tag{9}$$

Furthermore, (9) also yields $d_{\mathrm{TV}}\bigl(\mathscr{L}(S_n), \mathscr{L}(Y_n)\bigr) = \mathrm{o}(1)$. To see this, note that for any $a > 0$,

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(S_n), \mathscr{L}(Y_n)\bigr) \leqslant a d_n + \mathbb{P}[|S_n - \mu_n| \geqslant a] + \mathbb{P}[|Y_n - \mu_n| \geqslant a],$$

where d_n is the d_{loc}-bound from (9). The two probabilities can be estimated by Chebishev's inequality, yielding an order of $\mathrm{O}(\sigma_n^2 a^{-2})$. The choice

$$a = \left(\frac{\sigma_n^2}{d_n}\right)^{1/3}$$

solves $ad_n = \sigma_n^2 a^{-2}$, thus

$$d_{\mathrm{TV}}\big(\mathscr{L}(S_n), \mathscr{L}(Y_n)\big) = \mathrm{O}(ad_n + \sigma_n^2 a^{-2}) = \mathrm{O}\big(\sigma_n^{2/3} d_n^{2/3}\big) = \mathrm{o}(1). \tag{10}$$

Note, that the results in the papers of this thesis yield better rates that (10). In fact, in the examples of the presented papers, we will usually have $d_n = \mathrm{O}(\sigma_n^{-2})$, which, with the above result, yields $\mathrm{O}(\sigma_n^{-2/3})$, whereas with direct estimates we will obtain the expected order $\mathrm{O}(\sigma_n^{-1})$.

2.2. Translated Poisson approximation with Stein's method. I now show by means of the translated Poisson distribution the basic approach to obtaining approximations in terms of the above metrics. This approach is used in the first two papers of this thesis. It is very important to note that we are in a very different situation from that of the classical approximation results such as Le Cam (1960) and Barbour et al. (1992) for Poisson approximation. These results exploit the concept of *rare events*; that means, it is assumed that for a sum of (possibly dependent) indicator random variables the probability, say p_i, of each event i is small. There, one usually obtains rates of convergence of order $\mathrm{O}(\max p_i)$ to the Poisson distribution with a fixed mean.

We are however in the situation where $\max p_i$ does not converge to zero; for example, the binomial distribution with fixed p and $n \to \infty$. A good Poisson approximation cannot then be expected, simply because the quotient of mean and variance is bounded away from 1.

Now, the left side of (5), which is called the *Stein operator* for the standard normal distribution, can be generalised to the normal distribution with mean μ_n and variance σ_n^2, namely

$$\sigma_n^2 f'(x) - (x - \mu_n) f(x) \tag{11}$$

(changing of course the bounds in (7)). This suggests an equivalent operator on the integers, where we replace the first derivative $f'(j)$ by the first difference $\Delta f(j) = f(j+1) - f(j)$, thus

$$\sigma_n^2 \Delta f(j) - (j - \mu_n) f(j) \tag{12}$$

for $j \in \mathbb{Z}$, which turns out to be an appropriate Stein operator for the translated Poisson distribution $\mathrm{TP}(\mu_n, \sigma_n^2)$, that is, a Poisson distribution with mean σ_n^2 and shifted by $\mu_n - \sigma_n^2$ (in what follows, we omit the rounding problems; so assume that $\mu_n - \sigma_n^2 \in \mathbb{Z}$). Now, we replace the *Stein equation* (5) by

$$\sigma_n^2 \Delta f(j) - (j - \mu_n) f(j) = h(j) - \mathbb{E}h(Y_n), \tag{13}$$

where $\mathscr{L}(Y_n) \sim \mathrm{TP}(\mu, \sigma^2)$. We then find that the solution to (13) satisfies

$$\|f\|_\infty \leqslant \sigma_n^{-1}, \quad \|\Delta f\|_\infty \leqslant \sigma_n^{-2}, \quad \text{for all } h \in \mathcal{F}_{\mathrm{TV}} \tag{14}$$

(where we here in fact only use the bound on $\|\Delta f\|_\infty$). This will turn out to be enough for approximations in total variation. For the test functions $\mathcal{F}_{\mathrm{loc}}$ we have the better bound

$$\|f\|_\infty \leqslant \sigma_n^{-2}, \quad \text{for all } h \in \mathcal{F}_{\mathrm{loc}}. \tag{15}$$

I show now in case of the approximation of the binomial by a translated Poisson distribution, how the basic approach works, that is, how we estimate the l.h.s. of

$$\mathbb{E}\big\{\sigma_n^2\Delta f(S_n) - (S_n-\mu_n)f(S_n)\big\} = \mathbb{E}h(S_n) - \mathbb{E}h(Y_n). \tag{16}$$

Thus, let ξ_i, $i=1,\dots,n$ be i.i.d. random indicators with expectation p and S_n as before. One can construct a twice differentiable interpolation function $F:\mathbb{R}\to\mathbb{R}$ such that

$$F(j) = f(j), \qquad F'(j) = \Delta f(j)$$

for all $j\in\mathbb{Z}$. Thus we can replace the equality (16) by

$$\mathbb{E}\big\{\sigma_n^2F'(S_n) - (S_n-\mu_n)F(S_n)\big\} = \mathbb{E}h(S_n) - \mathbb{E}h(Y_n). \tag{17}$$

Recall that $\mu_n = np$ and $\sigma_n^2 = np(1-p)$, we thus have

$$\mathbb{E}\big\{\sigma_n^2F'(S_n) - (S_n-\mu_n)F(S_n)\big\} = \sum_{i=1}^n \mathbb{E}\big\{p(1-p)F'(S_n) - (\xi_i-p)F(S_n)\big\}. \tag{18}$$

Note now that, by Taylor expansion,

$$p(1-p)F'(S_n) = p(1-p)F'(S_n^i+p) + R_{i,1}, \tag{19}$$
$$(\xi-p)F(S_n) = (\xi-p)F(S_n^i+p) + (\xi_n-p)^2F'(S_n^i+p) + R_{i,2}, \tag{20}$$

where $S_n^i = S_n - \xi_i$ and

$$R_{i,1} = p(1-p)(\xi_n-p)\int_0^1 F''\big(S_n^i+p+s(\xi_i-p)\big)ds,$$

$$R_{i,2} = (\xi_i-p)^3\int_0^1 (1-s)F''\big(S_n^i+p+s(\xi_i-p)\big)ds$$

Now, $\mathbb{E}(\xi_i-p)=0$ and ξ_i and S_n^i are independent and therefore, putting (20) and (19) into the r.h.s. of (18), all terms except the remainder terms cancel, and we finally obtain

$$n\mathbb{E}\big\{R_{1,1}+R_{1,2}\big\} = \mathbb{E}h(S_n) - \mathbb{E}h(Y_n) \tag{21}$$

A naive estimate, for example for $R_{1,2}$, would yield

$$|\mathbb{E}R_{1,2}| \leqslant \tfrac12\|F''\|\mathbb{E}|\xi_1-p|^3 \leqslant \frac{C_1\mathbb{E}|\xi_1-p|^3}{2\sigma_n^2}$$

for an absolute constant C_1 (which comes from the interpolation), where the estimate $\|F''\|\leqslant C_1\sigma_n^{-2}$ cannot be improved. This is however not enough, as we would then obtain the final estimate

$$\big|\mathbb{E}h(S_n) - \mathbb{E}h(Z)\big| \leqslant C_1\frac{np(1-p)\mathbb{E}|\xi_1-p| + \frac12 n\mathbb{E}|\xi_1-p|^3}{np(1-p)} = \mathrm{O}(1)$$

for all $h \in \mathcal{F}_{\mathrm{TV}}$, which is of no use, even when p is small. So we must do better.

The key inequality is the following. Assume that one has given an integer valued random variable U and a bounded function g on $\mathbb{Z}$. Then

$$\mathbb{E}\Delta g(U) = \sum_{k\in\mathbb{Z}} p_k\bigl(g(k+1) - g(k)\bigr) = \sum_{k\in\mathbb{Z}} (p_{k-1} - p_k)g(k).$$

Thus

$$\bigl|\mathbb{E}\Delta g(U)\bigr| \leqslant \|g\|_\infty \sum_{k\in\mathbb{Z}} |p_{k-1} - p_k| = 2\|g\|_\infty\, d_{\mathrm{TV}}\bigl(\mathscr{L}(U), \mathscr{L}(U+1)\bigr). \tag{22}$$

Note now, that within the integral of the remainder terms we have more or less this situation: take F' as g, thus $F'' \approx \Delta g$, and S_n^i as U. In fact, recalling that $F' \approx \Delta f$, we can show that, for any z,

$$\bigl|\mathbb{E}F''(S_n^1 + z)\bigr| \leqslant 2C_2\|\Delta f\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(S_n^1), \mathscr{L}(S_n^1 + 1)\bigr) \tag{23}$$

for an absolute constant C_2 (which again comes from the interpolation). Thus, together with (14), we obtain the final estimate

$$\bigl|\mathbb{E}h(S_n) - \mathbb{E}h(Z_n)\bigr| \leqslant 2C_2 d_n \frac{np(1-p)\mathbb{E}|\xi_1 - p| + \frac{1}{2}n\mathbb{E}|\xi_1 - p|^3}{np(1-p)},$$

where $d_n = d_{\mathrm{TV}}\bigl(\mathscr{L}(S_n^1), \mathscr{L}(S_n^1 + 1)\bigr)$; this is almost the same as before, but with the additional factor d_n. This factor saves the day, because a standard coupling argument yields $d_n = \mathrm{O}(n^{-1/2})$. Thus, we have established that

$$d_{\mathrm{TV}}\bigl(\mathrm{Bi}(n,p), \mathrm{TP}(\mu_n, \sigma_n^2)\bigr) = \mathrm{O}(n^{-1/2}).$$

To obtain the d_{loc}-bound, we modify (22). Assume that an additional integer valued random variable V is given, independent of U. Then, one can show that

$$\bigl|\mathbb{E}\Delta^2 g(U+V)\bigr| \leqslant 4\|g\|_\infty\, d_{\mathrm{TV}}\bigl(\mathscr{L}(U), \mathscr{L}(U+1)\bigr)\, d_{\mathrm{TV}}\bigl(\mathscr{L}(V), \mathscr{L}(V+1)\bigr), \tag{24}$$

where Δ^2 denotes the second difference. Recalling the situation in (23), we see that we can interpret S_n^1 as the sum of two independent random variables, say $S_n^{1,1}$ and $S_n^{1,2}$, with about half of the summands each. Thus

$$\bigl|\mathbb{E}F''(S_n^1 + z)\bigr| \leqslant 4C_3\|f\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(S_n^{1,1}), \mathscr{L}(S_n^{1,1} + 1)\bigr)\, d_{\mathrm{TV}}\bigl(\mathscr{L}(S_n^{1,2}), \mathscr{L}(S_n^{1,2} + 1)\bigr) \tag{25}$$

This, together with the bound (15), yields

$$\bigl|\mathbb{E}h(S_n) - \mathbb{E}h(Z)\bigr| \leqslant 4C_3 d_n^1 d_n^2 \frac{np(1-p)\mathbb{E}|\xi_1 - p| + \frac{1}{2}n\mathbb{E}|\xi_1 - p|^3}{np(1-p)}$$

which is of the right order $\mathrm{O}(n^{-1})$.

Note that the expression $d_{\mathrm{TV}}\bigl(\mathscr{L}(U), \mathscr{L}(U+1)\bigr)$ in (22) is the analogue of the concentration inequality (4).

3. More about dependent summands

3.1. m-dependent sequences and local dependence. A sequence of random variables $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ is called *m-dependent* if for all i, the random vectors $(\xi_1, \ldots, \xi_i)$ and $(\xi_{i+t}, \ldots, \xi_n)$ are independent if $t > m$. One of the first ingenious approaches to prove a CLT under this condition was the idea of Bernstein (1927). He represents the S_n as a sum $S'_n + T_n$, where S'_n is known to satisfy a CLT (for example it may be a sum of independent summands) and where $\operatorname{Var} T_n / \operatorname{Var} S_n \to 0$ as $n \to \infty$. He then proves that S_n also satisfies a CLT. This idea was also used by Hoeffding and Robbins (1948) (and many later authors) to obtain CLTs under more general assumptions on the summands. Petrov (1960) was the first to obtain rates of convergence in the Kolmogorov metric, improved by Ibragimov (1967) to the optimal rate of $\mathrm{O}(n^{-(s-2)/2})$ under assumptions on the existence of the s-th moments of the summands where $2 < s \leqslant 3$. There are many other papers discussing m-dependence.

A possible generalisation is the concept of *local dependence*. In the context of the first two papers of this thesis, we call a family of random variables $\xi = (\xi_i)_{i \in I}$, where I is an arbitrary index set, *locally dependent*, if, for every i, there are sets $A_i \subset B_i \subset I$ such that ξ_i is independent of $(\xi_j)_{j \in A_i^c}$ and $(\xi_j)_{j \in A_i}$ is independent of $(\xi_j)_{j \in B_i^c}$; the set A_i is called the *neighbourhood* of i.

Chen and Shao (2004) give many results obtained with Stein's method with respect to the Kolmogorov-metric under (also other forms of) local dependence. Barbour et al. (1989) obtain results with respect to a smooth metric under a variant of local dependence also considered in the first two papers of this thesis.

LLTs for m-dependent summands can be found in Götze and Hipp (1990) and in Kazanchyan (2004), both papers using the method of characteristic functions.

In the first two papers of this thesis, we approximate sums of locally dependent random variables by a translated Poisson distribution and a centred and symmetric binomial distribution, respectively. The Taylor expansion is similar to (18)–(19) but becomes more involved. This would be enough to show a CLT, as has been illustrated for example in Barbour et al. (1989). To obtain the stronger results for d_{TV} and d_{loc}, one again needs estimates of the form (23) and (25); however, the sums involved cannot be represented any more as sums of independent random variables.

Consider the following simple example. Let $N = \{0, 1, 2, \ldots, n-1\}$ for some n and let $(I_{i,j})_{i,j \in N}$ be a family of independent random indicators with expectation p. Define

$$\xi_{i,j} := \prod_{k=i}^{i+1} \prod_{l=j}^{j+1} I_{k,l},$$

where we treat the indices n and 0 as being equal. If we think of the $I_{i,j}$ as lying on the grid $\mathbb{Z}^2$ (small sized numbers in Figure 1), we put a new random variable (large and bold numbers) in each square, which is the product of the four surrounding $I_{i,j}$'s. It is easy to see that the $\xi_{i,j}$ are locally dependent. However, if we condition on the values of the $I_{i,j}$ that lie on the borders of the 3×3-boxes as shown in Figure 1, we see that, though the $\xi_{i,j}$ within the boxes are still dependent, they

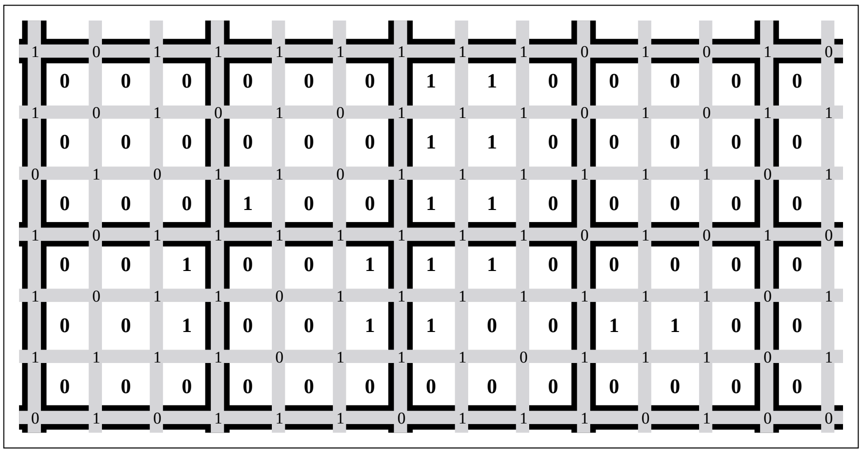

FIGURE 1. *The kilt problem. The small sized numbers on the grid are i.i.d. random indicators with expectation p and the larger numbers in the squares are the product of the four surrounding indicators. Though these larger numbers depend locally on each other, the blocks become independent if we condition on the indicators on the shaded subgrid.*

now become independent if they are in different boxes. Thus, S_n, conditioned on all the borders of the boxes, is a sum of independent integer valued random variables and this allows us to apply standard techniques to obtain the desired smoothing properties of S_n and thus to obtain bounds of the form (23) and (25).

3.2. Global dependence. There are many examples which do not fit into the scheme of local dependence. Consider for example the simple urn model corresponding to the hypergeometric distribution $\mathrm{Hyp}(n, m, N)$; that is, put m balls uniformly into N urns in such a way that there is at most one ball in each urn, and examine the number of balls in the first n urns. Note that the labelling of the urns in this example is irrelevant; that is, under any permutation of the urns, the resulting distribution is the same (though, of course, the concrete realisation will change). Thus the concept of 'neighbourhood' cannot be applied; instead, all urns depend equally on each other. We may call such a dependence *weak global dependence* in contrast to the local dependence. However, there are of course many examples between these two extremes, probably the most important being sequences and random fields satisfying some mixing conditions; see for example Doukhan (1994). For CLTs and LLTs on urn models see Kolchin et al. (1978).

In the framework of Stein's method, dependence of this type is typically handled with couplings. One of these couplings is called *exchangeable pair coupling*, and was introduced by Stein (1986). Besides the original random variable S_n, one constructs another random variable S_n' on the same probability space such that (S_n, S_n') and (S_n', S_n) have the same distribution. The construction of the coupling depends very much on the problem at hand; however, one general way has been proposed by Rinott and Rotar (1997). Construct a reversible discrete time Markov chain X such that its equilibrium distribution is equal to the distribution of S_n. Then, if the chain is in equilibrium, two consecutive steps will be an exchangeable pair. The need of reversibility can be weakend in some situations.

I illustrate these ideas with the anti-voter model, examined in the third paper of this thesis. For the following considerations, however, instead of a discrete time Markov chain we will construct a Markov jump process, with jump rates independent of the state of the process, and instead of the coupling we will study the generators of the processes involved. One can show that the coupling approach through the Markov chain and the generator approach through the Markov process sketched below are equivalent.

Assume that an n-vertex, r-regular graph G_n is given and assume that at every vertex there is a 'voter' having an opinion, either 0 or 1. Assume that the process is given at time point t. Wait an exponentially distributed rate $n/2$ amount of time and choose then uniformly a voter, say J. Choose uniformly a neighbour of voter J, say K. Then, assign to voter J the opinion opposite to that of voter K. We are now interested in the distribution of S_n, the number of voters having opinion 1 under the stationary distribution. Though the formulation of this problem is rather simple, the calculation of the stationary distribution turns out to be difficult; see Matloff (1977) and Donnelly and Welsh (1984). In the case of the anti-voter model, we could obtain the exchangeable pair for free: Assume that the process is in its stationary distribution; then the above transition mechanism already describes the coupling by taking the states immediately before and after a jump (note that although this Markov process is not reversible, one can show that the resulting pair is still exchangeable). However, as mentioned before, we will study the generators instead.

Recall that, to use Stein's method for the translated Poisson distribution, we have to bound the left side of (16). But before doing this, go back to the Stein operator of the normal distribution (11). We can in fact interpret this operator as the generator of a Markov diffusion process (just replace the function f by its derivative f' and the first derivative f' by the second derivative f''), in this case the so called Ornstein-Uhlenbeck process (scaled by σ_n and shifted by μ_n). Now, by bounding the Stein operator (16) we actually compare the dynamics of the above constructed process with the dynamics of an Ornstein-Uhlenbeck process; if the dynamics are similar, then also the two equilibrium distributions will be similar.

In the discrete setting, we replace the Ornstein-Uhlenbeck process by an immigration death process with per capita death rate 1 and immigration rate σ_n^2 and shift this process by $\mu_n - \sigma_n^2$. This process has the generator (12) if we replace f by Δf and Δf by $\Delta^2 f$. So we can see that both processes have a linear drift to the centre μ_n and a diffusion rate σ_n^2.

All we have to do now is to check whether the anti-voter process has similar dynamics. To this end we could calculate the generator of the anti-voter process,

$$(\mathcal{A}g)(k) = \lim_{h\to 0} \frac{\mathbb{E}\big\{g(S_n(t+h)) - g(k) \;\big|\; S_n(t) = k\big\}}{h}. \tag{26}$$

It turns out, however, that it is more convenient to calculate a more general object; instead of conditioning on $S_n(t)$, we condition on the all the individual voters $\xi(t)$

of $S_n(t)$. In this way, for every $x \in \{0,1\}^n$ and with $k = \sum_{i=1}^n x_i$, some calculation leads to

$$\begin{aligned}
(\tilde{\mathcal{A}}g)(x) &= \lim_{h\to 0} \frac{\mathbb{E}\big\{g(S_n(t+h)) - g(k) \,\big|\, \xi(t) = x\big\}}{h} \\
&= -\Big(k - \frac{n}{2}\Big)\Delta g(k-1) + \frac{1}{2r}\sum_{i=1}^n \sum_{j\in\mathcal{N}_i} (1-x_i)(1-x_j)\Delta^2 g(k-1) \qquad (27) \\
&= -\Big(k - \frac{n}{2}\Big) f(k) + \frac{1}{2r}\sum_{i=1}^n \sum_{j\in\mathcal{N}_i} (1-x_i)(1-x_j)\Delta f(k)
\end{aligned}$$

where $f(k) = \Delta g(k-1)$. From this we see that the anti-voter process has also a linear drift to the mean $\mu_n = n/2$, but a non-constant diffusion rate. However, if this rate does not fluctuate too much around σ_n^2, we expect a good approximation by a translated Poisson distribution. See Theorem 2.1 in the third paper for a rigorous statement. The key element in obtaining a generator of the simple form (27), without any remainder terms containing higher order differences, is the property that the exchangeable pair coupling satisfies

$$S_n' - S_n \in \{-1, 0, 1\}$$

almost surely.

References

W. J. Adams (1974). *The life and times of the central limit theorem.* Kaedmon Publishing Co., New York.

A. D. Barbour, Michał Karoński, and Andrzej Ruciński (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47**, 125–145.

A. D. Barbour, L. Holst, and S. Janson (1992). *Poisson approximation.* Oxford University Press, New York.

H. Bergström (1944). On the central limit theorem. *Skand. Aktuarietidskr.* **27**, 139–153.

S. Bernstein (1927). Sur l'extension du théoréme limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.* **97**, 1–59.

A. C. Berry (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Am. Math. Soc* **49**, 122–136.

E. Bolthausen (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**, 379–386.

P. L. Butzer and D. Schulz (1985). An extension of the Lindeberg-Trotter operator-theoretic approach to limit theorems for dependent random variables. I. General convergence theorems; approximation theorems with o-rates. *Acta Sci. Math. (Szeged)* **48**, 37–54.

L. Le Cam (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10**, 1181–1197.

L. H. Y. Chen and Q.-M. Shao (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, 1985–2028.

R. L. Dobrushin and B. Tirozzi (1977). The central limit theorem and the problem of equivalence of ensembles. *Comm. Math. Phys.* **54**, 173–192.

P. Donnelly and D. Welsh (1984). The antivoter problem: random 2-colourings of graphs. In *Graph theory and combinatorics (Cambridge, 1983)*, 133–144. Academic Press, London.

P. Doukhan (1994). *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

C. G. Esseen (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Math., Astro. Fysik* **28A**, 1–19.

B. V. Gnedenko (1948) On a local limit theorem of the theory of probability. *Uspehi Matem. Nauk (N. S.)* **3**, 187–194.

F. Götze and C. Hipp (1990). Local limit theorems for sums of finite range potentials of a Gibbsian random field. *Ann. Probab.* **18**, 810–828.

L. Heinrich (1982). A method for the derivation of limit theorems for sums of m-dependent random variables. *Z. Wahrsch. Verw. Gebiete* **60**, 501–515.

W. Hoeffding and H. Robbins (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773–780.

I. A. Ibragimov (1967). The central limit theorem for sums of functions of independent variables and sums of type $\sum f(2^k t)$. *Teor. Verojatnost. i Primenen* **12**, 655–665. (English translation in *Theor. Probability Appl.* **12** (1967), 596–607)

T. P. Kazanchyan (2004). A local limit theorem for sequences of dependent random variables. *Izv. Nats. Akad. Nauk Armenii Mat.* **39**, 33–42.

A. Y. Khinchin (1949). *Mathematical foundations of statistical mechanics.* Graylock Press, Albany.

A. Y. Khinchin (1960). *Mathematical foundations of quantum statistics.* Graylock Press, Albany.

V. F. Kolchin, B. A. Sevast′yanov, and V. P. Chistyakov (1978). *Random allocations.* V. H. Winston & Sons, Washington, D.C..

A. N. Kolmogorov (1949). A local limit theorem for classical Markov chains. *Izvestiya Akad. Nauk SSSR. Ser. Mat.* **13**, 281–300.

A. M. Liapunov (1900). Sur une proposition de la théorie des probabilités. *Bull. Aca. Sci. St. Petersbourg (5)* **13**, 359–386.

A. M. Liapunov (1901). Nouvelle forme du théoreme sur la limite de théorie des probabilités. *Mem. Aca. Sci. St. Petersbourg (8)* **12**, 1–24.

J. W. Lindeberg (1920). Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fenn., Ser. A* **16**, 1–23.

J. W. Lindeberg (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechung. *Math. Z.* **15**, 211–225.

N. S. Matloff (1977). Ergodicity conditions for a dissonant voting model. *Ann. Probability* **5**, 371–386.

L. V. Osipov and V. I. Rotar (1984). An infinite-dimensional central limit theorem. *Teor. Veroyatnost. i Primenen* **29**, 366–373. (English translation in *Theory Probab. Appl.* **29** (1984), 375–383)

V. V. Petrov (1960). On the central limit theorem for m-dependent quantities. In *Proc. All-Union Conf. Theory Prob. and Math. Statist. (Russian)*, 38–44. Izdat. Akad. Nauk Armjan. SSR, Erevan.

V. V. Petrov (1975). *Sums of independent random variables.* Springer-Verlag, New York.

Yu. V. Prokhorov and V. Statulevičius, editors (2000). *Limit theorems of probability theory.* Springer-Verlag, Berlin.

A. Račkauskas (1991). On the convergence rate in martingale CLT in Hilbert spaces. *Litovsk. Mat. Sb.* **31**, 497–512. (English translation in *Lithuanian Math. J.* **31** (1991), 345–355)

Y. Rinott and V. Rotar (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7**, 1080–1105.

Yu. A. Rozanov (1957). On a local limit theorem for lattice distributions. *Teor. Veroyatnost. i Primenen.* **2**, 275–281.

L. Saulis and V. A. Statulevičius (1991). Limit theorems for large deviations. In *Limit theorems of probability theory*, 185–266, Springer-Verlag, Berlin.

V. V. Sazonov (1981). *Normal approximation—some recent advances*, *Lecture Notes in Mathematics.* Springer-Verlag, Berlin.

V. Statuljavičjus (1961). Local limit theorems and asymptotic expansions for non-stationary Markov chains. *Litovsk. Mat. Sb.*, **1**, 231–314.

C. Stein (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Sympos. Math. Statist. Probability* **2**, 583–602.

C. Stein (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward.

A. N. Tihomirov (1980). Convergence rate in the central limit theorem for weakly dependent random variables. *Teor. Veroyatnost. i Primenen.* **25**, 800–818. (English translation in *Theory Probab. Appl.* **25** (1980), 790–890.)

H. F. Trotter (1959). An elementary proof of the central limit theorem. *Arch. Math.* **10**, 226–234.

published in *Bernoulli* **11** (2005), 1115–1128

APPROXIMATION OF SUMS OF CONDITIONALLY INDEPENDENT VARIABLES BY THE TRANSLATED POISSON DISTRIBUTION

By Adrian Röllin*

It is shown that the sum of a Poisson and an independent approximately normally distributed integer valued random variable can be well approximated in total variation by a translated Poisson distribution, and further that a mixed translated Poisson distribution is close to a mixed translated Poisson distribution with the same random shift but fixed variance. Using these two results, a general approach is then presented for the approximation of sums of integer valued random variables, having some conditional independence structure, by a translated Poisson distribution. We illustrate the method by means of two examples. The proofs are mainly based on Stein's method for distributional approximation.

1. Introduction

The Berry–Esseen theorem provides a uniform bound for the accuracy of the central limit theorem when approximating the probabilities of sets A of the form $(-\infty, a)$, $a \in \mathbb{R}$. If more complicated sets A are to be considered, some additional 'smoothness' condition is typically required. McDonald (1979) and Burgess and McDonald (1995) assumed a so-called 'Bernoulli part' to deduce a local limit theorem from a central limit theorem. Čekanavičius and Vaĭtkus (2001) used the smoothing property of a sum of independent Bernoulli random variables to approximate this sum with a translated Poisson distribution in total variation. Barbour and Čekanavičius (2002) incorporate a measure of the smoothness of the distribution of the individual independent integer-valued summands as a component of their estimate of the distance between the distribution of their sum and a translated Poisson distribution; see the discusion in the next section.

This paper combines ideas from the above papers to show that the distribution of many sums of dependent integer-valued random variables can be approximated in total variation by the translated Poisson distribution with the same order of accuracy as that of the Berry–Esseen theorem. Previous attempts are limited to simple examples (Barbour and Xia, 1999; Čekanavičius and Vaĭtkus, 2001). Analogous results hold also for local limit approximations.

Much in the spirit of McDonald (1979), we begin by considering the sum of an integer-valued random variable Φ, which is close in distribution to the normal, and an independent Poisson random variable, which acts as the smoothing component. We show that this sum can be well approximated in total variation by a translated Poisson distribution with the same first two moments (Theorem 1) and that a similar approximation follows for a local limit metric. The translated Poisson distribution, being concentrated on the integers, is a more natural approximation

*Partially supported by Schweizerischer Nationalfondsprojekt 20-107935/1

than the normal in the context of integer-valued variables, and the stronger results are a reflection of this.

We then show that Theorem 1 can be used as the basis of a rather general method, which yields good results in a number of dependent settings; see Theorem 3. We illustrate the method with two examples. For the proofs, we use Stein's method for distributional approximation, introduced by Stein (1972), adapted to the Poisson setting; see (Barbour et al., 1992).

1.1. Notation. We say that an integer-valued random variable Y has a *translated Poisson distribution* with parameters μ and σ^2 and write

$$\mathscr{L}(Y) = \mathrm{TP}(\mu, \sigma^2)$$

if $\mathscr{L}(Y - \mu + \sigma^2 + \gamma) = \mathrm{Po}(\sigma^2 + \gamma)$ where $\gamma = \langle \mu - \sigma^2 \rangle$ and $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that $\mathbb{E}Y = \mu$ and that $\sigma^2 \leqslant \operatorname{Var} Y = \sigma^2 + \gamma \leqslant \sigma^2 + 1$. Note also that $\mathrm{Po}(\sigma^2) = \mathrm{TP}(\sigma^2, \sigma^2)$.

We say that an integer-valued random variable Y has an *F-mixed translated Poisson distribution* and write

$$\mathscr{L}(Y) = \mathrm{TP}[F]$$

if F is a probability measure on $\mathbb{R} \times \mathbb{R}^+$ and, for all $j \in \mathbb{Z}$,

$$\mathbb{P}[Y = j] = \int\limits_{\mathbb{R}\times\mathbb{R}^+} \mathrm{TP}(x, y)\{j\}\, F(dx, dy).$$

Thus a mixed Poisson distribution $\mathrm{Po}[G]$ with mixing distribution G is $\mathrm{TP}[F]$, where F is concentrated on the diagonal and has marginals G.

In this paper, the measure F will often be generated by two random variables Φ and Λ on a common probability space, that is, $F := \mathscr{L}(\Phi, \Lambda)$. We treat Φ as the 'random shift' and Λ as the 'random variance' of Y. Note that, due to our definition of $\mathrm{TP}(\mu, \sigma^2)$, Φ need not be integer-valued.

Throughout the paper, we shall be concerned with two metrics for probability distributions on the integers, the total variation metric d_{TV} and the local limit metric d_{loc}, where for two probability distributions P and Q,

$$d_{\mathrm{TV}}(P, Q) := \sup_{A\subset\mathbb{Z}} \bigl|P(A) - Q(A)\bigr|, \qquad d_{\mathrm{loc}}(P, Q) := \sup_{k\in\mathbb{Z}} \bigl|P(\{k\}) - Q(\{k\})\bigr|.$$

2. Main results

2.1. Poisson smoothing. In this paper, we assume the random translation Φ to be approximately Gaussian. In terms of Stein's method of distributional approximation, this is to be expressed as follows. Denote by $\|\cdot\|$ the essential supremum norm and define the function space

$$\mathcal{F} = \bigl\{ f \in C^1(\mathbb{R}) \bigm| f' \text{ absolutely continuous}, \|f\| + \|f'\| + \|f''\| < \infty \bigr\}.$$

Then, we shall assume that, for some $\varepsilon \geqslant 0$,

$$\bigl|\mathbb{E}\bigl\{f'(\Phi_c) - \Phi_c f(\Phi_c)\bigr\}\bigr| \leqslant \varepsilon\|f''\|, \qquad \text{for all } f \in \mathcal{F}, \tag{2.1}$$

where $\Phi_c := (\Phi - \mu)/\tau$ and μ and τ^2 are the mean and variance of Φ.

Theorem 1. *Let Φ be a random variable with mean μ and variance τ^2 such that estimate* (2.1) *holds for some $\varepsilon \geqslant 0$. Then, for any $\lambda > 0$,*

$$d_{\mathrm{TV}}\bigl(\mathrm{TP}[\mathscr{L}(\Phi) \times \delta_\lambda], \mathrm{TP}(\mu, \tau^2 + \lambda)\bigr) \leqslant \frac{c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 3\sqrt{\lambda}}{(\tau^2 + \lambda)\sqrt{\lambda}}, \tag{2.2}$$

$$d_{\mathrm{loc}}\bigl(\mathrm{TP}[\mathscr{L}(\Phi) \times \delta_\lambda], \mathrm{TP}(\mu, \tau^2 + \lambda)\bigr) \leqslant \frac{4c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 12\sqrt{\lambda}}{(\tau^2 + \lambda)\lambda}, \tag{2.3}$$

where $c_0 = 1 + \sqrt{2}$.

So, suppose that $(\Phi^{(n)})_{n\geqslant 1}$ is a sequence obeying a central limit theorem, in the sense that $\Phi_c^{(n)}$ converges to the standard normal and that the corresponding sequence $(\varepsilon_n)_{n\geqslant 1}$ from (2.1) tends to zero. Suppose also that $\tau_n^2 := \operatorname{Var}\Phi^{(n)}$ and λ_n tend to infinity at the same rate as $n \to \infty$. Then the estimate (2.2) is of order $\mathrm{O}\bigl(\varepsilon_n + \lambda_n^{-1/2}\bigr)$ and (2.3) is of order $\mathrm{O}\bigl(\lambda_n^{-1/2}\varepsilon_n + \lambda_n^{-1}\bigr)$. In typical situations, say $\tau_n^2 \asymp n$ in a central limit theorem for sums of locally dependent variables, we recover the expected order $\mathrm{O}(n^{-1/2})$ for (2.2) and $\mathrm{O}(n^{-1})$ for (2.3) if $\lambda_n \asymp n$; compare these with the second example in the next section.

2.2. Translated Poisson approximation. Let W be an integer-valued random variable with mean μ and variance σ^2 and X a random element of a Polish space on the same probability space. Assume that we want to approximate $\mathscr{L}(W)$ by a translated Poisson distribution with parameters μ and σ^2. Put $\mu_X = \mathbb{E}(W|X)$, $\sigma_X^2 = \operatorname{Var}(W|X)$ and $\lambda = \mathbb{E}(\sigma_X^2)$ and consider the following, simple application of the triangle inequality for a metric d:

$$\begin{aligned} d\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant{} & d\bigl(\mathscr{L}(W), \mathrm{TP}[\mathscr{L}(\mu_X, \sigma_X^2)]\bigr) \\ & + d\bigl(\mathrm{TP}[\mathscr{L}(\mu_X, \sigma_X^2)], \mathrm{TP}[\mathscr{L}(\mu_X) \times \delta_\lambda]\bigr) \\ & + d\bigl(\mathrm{TP}[\mathscr{L}(\mu_X) \times \delta_\lambda], \mathrm{TP}(\mu, \sigma^2)\bigr), \end{aligned} \tag{2.4}$$

where in this paper we shall take either d_{TV} or d_{loc}.

The second term on the right in (2.4) can be bounded using Stein's method, as in the next theorem.

Theorem 2. *Let Φ be a real-valued random variable and let Λ be a non-negative random variable with expectation $\lambda > 0$ and variance ν^2. Then*

$$d_{\mathrm{TV}}\bigl(\mathrm{TP}[\mathscr{L}(\Phi, \Lambda)], \mathrm{TP}[\mathscr{L}(\Phi) \times \delta_\lambda]\bigr) \leqslant \frac{2 + \nu}{\lambda} + \frac{1}{\lambda^2}, \tag{2.5}$$

$$d_{\mathrm{loc}}\bigl(\mathrm{TP}[\mathscr{L}(\Phi, \Lambda)], \mathrm{TP}[\mathscr{L}(\Phi) \times \delta_\lambda]\bigr) \leqslant \frac{2\sqrt{2}(1 + \nu) + \sqrt{\lambda}}{\lambda^{3/2}} + \frac{1 + 4\nu^2}{\lambda^2}. \tag{2.6}$$

The bounds (2.2)–(2.3) and (2.5)–(2.6) will be used for large λ, and typically with τ^2 and ν^2 large as well. It is, however, interesting to note that they do not tend to 0 if τ^2 and ν^2 tend to 0, as might have been expected. The reason is that the distributions $\mathrm{TP}(\mu, \sigma^2)$, although indexed by two continuous parameters, all belong to the set $\delta_m * \mathrm{Po}(\lambda)$ for $(m, \lambda) \in \mathbb{Z} \times \mathbb{R}_+$, where δ_m denotes the unit mass at m and $*$ the convolution of measures. This is reflected by the fact that the distributions $\mathrm{TP}(\mu, \sigma^2)$ do not change continuously with respect to either μ or σ^2 when $\mu - \sigma^2 \in \mathbb{Z}$. For example, $\mathrm{TP}(2-\varepsilon, 1) = \mathrm{Po}(2-\varepsilon)$, but $\mathrm{TP}(2,1) = 1 + \mathrm{Po}(1)$. Because of this fundamental discontinuity, $\tau^2 \to 0$ and $\nu^2 \to 0$ cannot imply that the bounds (2.2)–(2.3) and (2.5)–(2.6) tend to zero.

Barbour et al. (1992) gave a bound for the distance $d_{\mathrm{TV}}\bigl(\mathrm{Po}[\mathscr{L}(\Lambda)], \mathrm{Po}(\lambda)\bigr)$ of order $\mathrm{O}(\nu^2/\lambda)$. However, $\mathscr{L}(\Lambda)$ influences both the mean and variance of the distribution $\mathrm{Po}[\mathscr{L}(\Lambda)]$, whereas in Theorem 2 it only mixes the variance of $\mathrm{TP}[\mathscr{L}(\Phi, \Lambda)]$, leading to qualitatively different bounds.

To bound the third term on the right of (2.4), we can apply Theorem 1, provided that μ_X satisfies inequality (2.1) for some small ε. Combining all the above facts, we have the following theorem.

Theorem 3. *Let W be an integer-valued random variable with expectation μ and variance σ^2 and let X be a random element of a Polish space on the same probability space. Define μ_X, σ_X^2 and λ as at the beginning of this section and let $\tau^2 = \mathrm{Var}(\mu_X)$, $\nu^2 = \mathrm{Var}(\sigma_X^2)$. Assume that there exists $\varepsilon \geqslant 0$ such that $(\mu_X - \mu)/\tau$ satisfies (2.1). Then*

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant \mathbb{E}D_{\mathrm{TV}}(X) + \frac{2+\nu}{\lambda} + \frac{1}{\lambda^2} + \frac{c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 3\sqrt{\lambda}}{\sigma^2\sqrt{\lambda}},$$

$$d_{\mathrm{loc}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant \mathbb{E}D_{\mathrm{loc}}(X) + \frac{2\sqrt{2}(1+\nu) + \sqrt{\lambda}}{\lambda^{3/2}} + \frac{1+4\nu^2}{\lambda^2} + \frac{4c_0(2\varepsilon\tau^3 + 2\tau^2 + \tau) + 12\sqrt{\lambda}}{\sigma^2\lambda},$$

where $c_0 = 1 + \sqrt{2}$ and

$$D_{\mathrm{TV}}(X) := d_{\mathrm{TV}}\bigl(\mathscr{L}(W|X), \mathrm{TP}(\mu_X, \sigma_X^2)\bigr),$$
$$D_{\mathrm{loc}}(X) := d_{\mathrm{loc}}\bigl(\mathscr{L}(W|X), \mathrm{TP}(\mu_X, \sigma_X^2)\bigr).$$

Now we are already able to bound $\mathbb{E}D_{\mathrm{TV}}(X)$ and $\mathbb{E}D_{\mathrm{loc}}(X)$ if the conditional distribution $\mathscr{L}(W|X)$ can be represented as a sum of independent integer random variables, since, as in Barbour and Čekanavičius (2002) or Čekanavičius and Vaĭtkus (2001), we can then approximate $\mathscr{L}(W|X)$ by the corresponding translated Poisson distribution.

To see this in more detail, recall Theorem 3.1 of Barbour and Čekanavičius (2002). Let $\widetilde{W} = \sum_{i=1}^n Z_i$ be a sum of independent integer-valued random variables, such that $\mathbb{E}Z_i = \mu_i$, $\operatorname{Var} Z_i = \sigma_i^2$ and $\mathbb{E}|Z_i^3| < \infty$. Put

$$\widetilde{W}_i := \widetilde{W} - Z_i, \qquad d := \max_{1 \leqslant i \leqslant n} d_{\mathrm{TV}}\big(\mathscr{L}(\widetilde{W}_i), \mathscr{L}(\widetilde{W}_i + 1)\big), \tag{2.7}$$

$$\psi_i := \sigma_i^2 \mathbb{E}\{Z_i(Z_i - 1)\} + |\mu_i - \sigma_i^2| \mathbb{E}\{(Z_i - 1)(Z_i - 2)\} + \mathbb{E}|Z_i(Z_i - 1)(Z_i - 2)|. \tag{2.8}$$

Then, with $\tilde{\mu} = \sum \mu_i$, $\tilde{\sigma}^2 = \sum \sigma_i^2$, and $\psi = \sum \psi_i$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(\widetilde{W}), \mathrm{TP}(\tilde{\mu}, \tilde{\sigma}^2)\big) \leqslant \frac{2 + d\psi}{\tilde{\sigma}^2}. \tag{2.9}$$

The factor d may be expressed in terms of the smoothness of the individual Z_i. With

$$v_i := \min\big\{\tfrac{1}{2}, 1 - d_{\mathrm{TV}}\big(\mathscr{L}(Z_i), \mathscr{L}(Z_i + 1)\big)\big\} \tag{2.10}$$

we have the simpler bound

$$d \leqslant \Big(\sum_{i=1}^{n} v_i - \max_{1 \leqslant i \leqslant n} v_i\Big)^{-1/2}. \tag{2.11}$$

For analogous bounds in the d_{loc} case, we need some further notation. Proceeding as Barbour and Čekanavičius (2002, Section 4), define

$$d' := \tfrac{1}{2} \max_{1 \leqslant i \leqslant n} \|\mathscr{L}(W_i) * (\delta_1 - \delta_0)^{*2}\|. \tag{2.12}$$

Using (4.4) and (4.8), just slight adaptations to the proof of Theorem 3.1 in (Barbour and Čekanavičius, 2002) are needed to show that

$$d_{\mathrm{loc}}\big(\mathscr{L}(\widetilde{W}), \mathrm{TP}(\tilde{\mu}, \tilde{\sigma}^2)\big) \leqslant \frac{2 + d'\psi}{\tilde{\sigma}^2}. \tag{2.13}$$

From equation (4.9) in (Barbour and Čekanavičius, 2002) we obtain the bound

$$d' \leqslant 4\Big(\sum_{i=1}^{n} v_i - 4 \max_{1 \leqslant i \leqslant n} v_i\Big)^{-1}. \tag{2.14}$$

3. Applications

3.1. Random sum of independent and identically distributed random variables.

Theorem 4. *Let N be a non-negative, integer-valued random variable with expectation $a > 8$ and variance b^2 such that* (2.1) *holds for $N_c := (N - a)/b$ and some $\varepsilon \geqslant 0$, and let $Z_1, Z_2, \ldots$ be independent and identically distributed integer-valued random variables with expectation r and variance s^2, independent also of N; put $W = \sum_{i=1}^{N} Z_i$. Let ψ_1 and v_1 be as in* (2.8) *and* (2.10) *for Z_1, and assume that*

$v_1 > 0$. *Then, with* $\mu = \mathbb{E}W = ar$ *and* $\sigma^2 = \operatorname{Var} W = as^2 + b^2r^2$,

$$
\begin{aligned}
&d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\big)\\
&\qquad \leqslant \frac{1.5\psi_1}{s^2\sqrt{v_1(a-2)}} + \frac{5\varepsilon b^3r^3 + 5b^2r^2 + 2.5br}{(as^2+b^2r^2)sa^{1/2}} + \frac{1 + 9as^2 + abs^4 + 4b^2s^4}{a^2s^4},\\
&d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\big)\\
&\qquad \leqslant \frac{8\psi_1}{s^2v_1(a-8)} + \frac{15+3bs^2}{a^{3/2}s^3} + \frac{1+5as^2+8b^2s^4}{a^2s^4} + \frac{10(2\varepsilon b^3r^2 + 2b^2r^2 + br)}{(as^2+b^2r^2)as^2}
\end{aligned}
$$

A random variable W of the form considered in this example arises in the study of the Reed–Frost epidemic process treated by Barbour and Utev (2004). In their Theorem 3.1, a local limit theorem is proved using Fourier arguments under the assumption that the Laplace transform of N is close to that of the normal distribution. Our result is formulated in very much simpler terms, and in addition gives an explicit approximation error. If we assume that $a \asymp n$ and $b^2 \asymp n$ and that $\varepsilon = \mathrm{O}(n^{-1/2})$, the total variation bound above is of order $\mathrm{O}(n^{-1/2})$.

Barbour and Utev (2004, Theorem 3.2), also prove a stronger local limit approximation, but at the cost of very much more restrictive conditions than ours.

Proof. We apply Theorem 3. In accordance with the notation of the previous section, let

$$
\begin{gathered}
\mu_N := \mathbb{E}(W|N) = Nr, \quad \tau^2 := \operatorname{Var}(\mu_N) = b^2r^2;\\
\sigma_N^2 := \operatorname{Var}(W|N) = Ns^2, \quad \lambda := \mathbb{E}(\sigma_N^2) = as^2, \quad \nu^2 := \operatorname{Var}(\sigma_N^2) = b^2s^4.
\end{gathered}
$$

Then, given $N = k$, we can apply Theorem 3.1 from Barbour and Čekanavičius (2002) to W in order to bound $D_{\mathrm{TV}}(k)$ and $D_{\mathrm{loc}}(k)$. To this end, define $d(k)$ as in (2.7) and $d'(k)$ as in (2.12) with $n = k$. From (2.11), we obtain the estimate $d(k) \leqslant (kv_1 - v_1)^{-1/2}$, and from (2.14), $d'(k) \leqslant 4(kv_1 - 4v_1)^{-1}$ and hence, applying (2.9) and (2.13),

$$
D_{\mathrm{TV}}(k) \leqslant \begin{cases} \dfrac{4}{as^2} + \dfrac{\psi_1\sqrt{2}}{s^2\sqrt{v_1(a-2)}} & \text{if } k \geqslant a/2,\\ 1 & \text{if } k < a/2, \end{cases}
$$

$$
D_{\mathrm{loc}}(k) \leqslant \begin{cases} \dfrac{4}{as^2} + \dfrac{8\psi_1}{s^2v_1(a-8)} & \text{if } k \geqslant a/2,\\ 1 & \text{if } k < a/2. \end{cases}
$$

Using Chebyshev's inequality to bound $\mathbb{P}[N < a/2]$, we therefore obtain

$$
\begin{aligned}
\mathbb{E}D_{\mathrm{TV}}(N) &\leqslant \frac{4b^2}{a^2} + \frac{4}{as^2} + \frac{\psi_1\sqrt{2}}{s^2\sqrt{v_1(a-2)}},\\
\mathbb{E}D_{\mathrm{loc}}(N) &\leqslant \frac{4b^2}{a^2} + \frac{4}{as^2} + \frac{8\psi_1}{s^2v_1(a-8)}.
\end{aligned}
$$

The remaining elements in Theorem 3 are immediate; we use $\sqrt{2} \leqslant 1.5$ and hence $c_0 \leqslant 2.5$. ∎

3.2. k-runs.

Theorem 5. *Let $\xi_0, \ldots, \xi_{n-1}$ be independent and identically distributed random variables with $\mathbb{P}[\xi_0 = 1] = 1 - \mathbb{P}[\xi_0 = 0] = p$ for some $p \in (0,1)$, where $n = m(2k-1)$ for some integers $k, m \geqslant 2$. To avoid edge effects, put $\xi_{n+i} := \xi_i$ for $i = 0, \ldots, 2k-2$. Define $U_j := \prod_{i=j}^{j+k-1} \xi_i$ and put $W = \sum_{j=0}^{n-1} U_j$. Then, with $\mu = \mathbb{E}W = np^k$ and $\sigma^2 = \operatorname{Var} W = np^k\bigl\{1get + p - p^k(2 + (2k-1)(1-p))\bigr\}/(1-p)$,*

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant \frac{K_1}{\sqrt{n}}, \qquad d_{\mathrm{loc}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant \frac{K_2}{n}$$

for some constants $K_i = K_i(k,p)$, $i = 1, 2$, which are independent of n.

The formulas for $K_i(k,p)$ that we establish here are rather crude and complicated but explicit. For $k = 2$, a bound of the same order was given by (Barbour and Xia, 1999), but their method of proof was extremely involved. Here, we can apply Theorem 3 and (2.9) rather directly, to obtain a result for arbitrary k. Some numerical comparisons with the bound of (Barbour and Xia, 1999) for $k = 2$ are given in Table 1, deduced by a more careful examination of the error terms.

Proof. Once again we apply Theorem 3. Split the indices $N_n := \{0, \ldots, n-1\}$ into m blocks $J_b = J_b^1 \cup J_b^2$, $b \in N_m$, of size $s := 2k-1$ with $J_b^1 = \bigl\{bs, \ldots, bs + k - 2\bigr\}$ and $J_b^2 = \bigl\{bs + k - 1, \ldots, (b+1)s - 1\bigr\}$, and set

$$X = \Bigl\{\xi_i \,:\, i \in \bigcup_{b \in N_m} J_b^1\Bigr\}.$$

Let L_b (R_b) be the number of consecutive 1s of the ξ_i at the beginning (end) of block J_b^1. With $W = \sum_{b \in N_m} W_b$, where $W_b = \sum_{j \in J_b} U_j$, the W_b's are conditionally independent given X. Note that $\mathscr{L}(W_b | X) = \mathscr{L}(W_1 | R_1, L_2)$. We have

$$\mathbb{E}\bigl\{W_1 \bigm| R_1 = r, L_2 = l\bigr\} = \frac{p^{k-r} + p^{k-l} - 2p^k}{1-p} + p^k$$

and so

$$\mu_X = mp^k + \sum_{b \in N_m} \frac{p^{k-R_b} + p^{k-L_{b+1}} - 2p^k}{1-p} = mp^k + \sum_{b \in N_m} V_b,$$

where the $V_b := (p^{k-R_b} + p^{k-L_b} - 2p^k)/(1-p)$ are independent and identically distributed with $\mathbb{E}V_1 = 2(k-1)p^k$. Some simple calculations give

$$\mathbb{E}\bigl\{p^{-2L_1}\bigr\} = \frac{1 + p - p^k}{p^{k-1}}, \quad \mathbb{E}\bigl\{p^{-L_1}\bigr\} = k - (k-1)p,$$
$$\mathbb{E}\bigl\{p^{-L_1-R_1}\bigr\} = p^{-k+1} + (k-1)(1-p) + \tfrac{1}{2}(k-1)(k-2)(1-p)^2,$$

hence

$$\begin{aligned}
&\tau_1^2(k,p) := \operatorname{Var} V_1 \\
&= \frac{p^{k+1}}{(1-p)^2}\bigl(4 + 2p - (3k^2 + k)p^{k-1} + (6k^2 - 4k - 4)p^k - (3k^2 - 5k + 2)p^{k+1}\bigr)
\end{aligned}$$

and

$$\tau^2 := \operatorname{Var}\mu_X = m\tau_1^2 \leqslant \frac{mp^{k+1}(4+2p)}{(1-p)^2} \tag{3.1}$$

As $\bigl|V_b - \mathbb{E}V_b\bigr| \leqslant 2p/(1-p)$ almost surely, we have $\mathbb{E}\bigl|V_b - \mathbb{E}V_b\bigr|^3 \leqslant 2p\tau_1^2/(1-p)$. Now, an inequality of the form (2.1) is easily derived, see for example Reinert (1998, Theorem 2.1): For a sum of independent random variables $\sum Z_i$ with zero expectation and variances σ_i^2 such that $\sum \sigma_i^2 = 1$, inequality (2.1) holds with $\varepsilon = \sum(\sigma_i^3 + \frac{1}{2}\mathbb{E}|Z_i^3|)$, and we may therefore take

$$\varepsilon = \frac{1}{\sqrt{m}}\biggl\{1 + \frac{2p}{(1-p)\tau_1}\biggr\} =: \frac{1}{\sqrt{m}}\varepsilon_1(k,p). \tag{3.2}$$

For (2.9), we have the following rather crude bounds. First, note that

$$\mathbb{P}\bigl[W_b = 0 \bigm| R_b, L_{b+1}\bigr] \geqslant (1-p)^2, \quad \mathbb{P}\bigl[W_b = 1 \bigm| R_b, L_{b+1}\bigr] \geqslant p^k(1-p)^2 \tag{3.3}$$

almost surely. Hence, from (3.3),

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W_b|R_b, L_{b+1}), \mathscr{L}(W_b+1|R_b, L_{b+1})\bigr) \leqslant 1 - p^k(1-p)^2$$

and with (2.11) and (2.14)

$$d \leqslant p^{-k/2}(1-p)^{-1}(m-1)^{-1/2}, \qquad d' \leqslant 4p^{-k}(1-p)^{-2}(m-4)^{-1} \tag{3.4}$$

Furthermore, it follows from (3.3) that

$$\operatorname{Var}(W_b|R_b, L_{b+1}) \geqslant p^k(1-p)^2, \tag{3.5}$$

and, noting that $0 \leqslant W_b \leqslant sI[W_b \geqslant 1]$,

$$\operatorname{Var}(W_b|R_b, L_{b+1}) \leqslant \mathbb{E}(W_b^2|R_b, L_{b+1}) \leqslant s^2 P_b,$$

where $P_b := \mathbb{P}[W_b \geqslant 1|R_b, L_{b+1}]$. Thus $\psi_b(R_b, L_{b+1}) \leqslant s^3P_b^2(1+2s) + s^3P_b$ and since

$$\mathbb{E}P_b^2 \leqslant \mathbb{E}P_b = \mathbb{P}[W_b \geqslant 1] \leqslant \mathbb{E}W_b = sp^k \tag{3.6}$$

it follows that

$$\mathbb{E}\psi_b(R_b, L_{b+1}) \leqslant 2p^k s^4(1+s). \tag{3.7}$$

Thus, from (2.9), (3.4), (3.5) and (3.7),

$$\mathbb{E}D_{\mathrm{TV}}(X) \leqslant \frac{2}{mp^k(1-p)^2} + \frac{4k(2k-1)^4}{(1-p)^3\sqrt{(m-1)p^k}}, \tag{3.8}$$

$$\mathbb{E}D_{\mathrm{loc}}(X) \leqslant \frac{2}{mp^k(1-p)^2} + \frac{16k(2k-1)^4}{p^k(1-p)^4(m-4)}. \tag{3.9}$$

To complete the bound in Theorem 3, we still need a lower bound for λ and an upper bound for ν^2, both of which are properties of the distribution of

$$\sigma_X^2 = \operatorname{Var}\biggl(\sum_{b=0}^{m-1} W_b \biggm| X\biggr) = \sum_{b=0}^{m-1}\operatorname{Var}(W_b|R_b, L_{b+1}) =: \sum_{b=0}^{m-1} Y_b.$$

(a)	0.1	0.25	0.5	0.75	0.9
10^6	0.4463	0.2334	0.1747	0.5528	>1
10^8	0.0445	0.0233	0.0175	0.0553	0.2554
10^{10}	0.0045	0.0023	0.0017	0.0055	0.0255

(b)	0.1	0.25	0.5	0.75	0.9
10^6	0.0304	—	0.1251	0.6014	—
10^8	0.0030	—	0.0125	0.0601	—
10^{10}	0.0003	—	0.0013	0.0060	—

TABLE 1
Numerical comparison of the 2-runs example: total variation distance estimate using the method in (a) this paper and (b) Barbour and Xia (1999). Missing values are due to parameter restrictions.

It is immediate from (3.3) that

$$\lambda = \mathbb{E}(\sigma_X^2) = m\mathbb{E}Y_1 \geqslant mp^k(1-p)^2, \tag{3.10}$$

and, since the Y_b are 1-dependent,

$$\nu^2 = \mathrm{Var}(\sigma_X^2) = \sum_{j=0}^{m-1} \mathrm{Var}\, Y_b + 2\sum_{j=0}^{m-1} \mathrm{Cov}(Y_b, Y_{b+1}) \leqslant 3m \,\mathrm{Var}\, Y_1.$$

Now $\mathrm{Var}\, Y_1 \leqslant \mathbb{E}Y_1^2$ and

$$Y_1 = \mathrm{Var}(W_1 | R_1, L_2) \leqslant \mathbb{E}(W_1^2 | R_1, L_2) \leqslant s^2 P_1$$

almost surely, so that, with (3.6), $\mathrm{Var}\, Y_1 \leqslant s^4\mathbb{E}(P_1^2) \leqslant p^k s^5$; hence

$$\nu^2 \leqslant 3mp^k(2k-1)^5. \tag{3.11}$$

Combining (3.1), (3.2), (3.8), (3.9), (3.10) and (3.11) with the bounds in Theorem 3 it follows that d_{TV} is of order $\mathrm{O}(m^{-1/2})$ and d_{loc} of order $\mathrm{O}(m^{-1})$ and recalling that $m = n/(2k-1)$ completes the proof. $\square$

4. Proofs

4.1. Stein approach for the translated Poisson distribution. To use Stein's method for approximation in the d_{TV} and d_{loc} metrics we start with the Poisson case; for details see Barbour et al. (1992).

Let W be an integer-valued random variable with expectation μ and variance $\sigma^2 > 0$, and let $s = \lfloor \mu - \sigma^2 \rfloor$ and $\gamma = \langle \mu - \sigma^2 \rangle$ where $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that, if $Y \sim \mathrm{TP}(\mu, \sigma^2)$, $Y - s \sim \mathrm{Po}(\sigma^2 + \gamma)$. Let $\mathcal{A}g(j) = (\sigma^2+\gamma)g(j+1) - jg(j)$ be the usual Stein operator for the Poisson distribution with mean $\sigma^2 + \gamma$, and for $A \subset \mathbb{Z}_+ := \{0, 1, 2, \dots\}$ let $g_A : \mathbb{Z} \to \mathbb{R}$ be the (bounded) solution of

i) $g(j) = 0$ for all $j \leqslant 0$,
ii) $\mathcal{A}g(j) = I[j \in A] - \mathrm{Po}(\sigma^2+\gamma)\{A\}$ for all $j > 0$.

We can thus bound the total variation distance with

$$\begin{aligned} d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) &= d_{\mathrm{TV}}\bigl(\mathscr{L}(W-s), \mathrm{Po}(\sigma^2+\gamma)\bigr) \\ &= \sup_{B\subset\mathbb{Z}}\bigl|\mathbb{E}I[W-s\in B] - \mathrm{Po}(\sigma^2+\gamma)\{B\}\bigr| \\ &\leqslant \sup_{A\subset\mathbb{Z}_+}\bigl|\mathbb{E}\mathcal{A}g_A(W-s)\bigr| + \mathbb{P}[W-s<0], \end{aligned} \tag{4.1}$$

and analogously

$$d_{\mathrm{loc}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) \leqslant \sup_{k\in\mathbb{Z}_+}\bigl|\mathbb{E}\mathcal{A}g_{\{k\}}(W-s)\bigr| + \mathbb{P}[W-s<0]. \tag{4.2}$$

The last terms in (4.1) and (4.2) are usually bounded using Chebyshev's inequality.

From Barbour et al. (1992) we obtain the well-known bounds on the supremum norm of g_A,

$$\|g_A\| \leqslant (\sigma^2+\gamma)^{-1/2} \leqslant \sigma^{-1}, \quad \|\Delta g_A\| \leqslant (\sigma^2+\gamma)^{-1} \leqslant \sigma^{-2}, \tag{4.3}$$

where $\Delta g_A(j) := g_A(j+1) - g_A(j)$. If $A = \{k\}$ for some $k \in \mathbb{Z}$, we have the better estimate

$$\|g_{\{k\}}\| \leqslant (\sigma^2+\gamma)^{-1} \leqslant \sigma^{-2}. \tag{4.4}$$

With $\tilde{g}_A(j) := g_A(j-s)$ we can rewrite the Stein operator, obtaining

$$\begin{aligned} \mathcal{A}g_A(W-s) &= (\sigma^2+\gamma)g_A(W-s+1) - (W-s)g_A(W-s) \\ &= \sigma^2\Delta\tilde{g}_A(W) - (W-\mu)\tilde{g}_A(W) + \gamma\Delta\tilde{g}_A(W). \end{aligned} \tag{4.5}$$

The bounds on $\tilde{g}_A$ are of course the same as on g_A in (4.3) and (4.4). Thus, the last term is easily bounded by

$$\bigl|\mathbb{E}\{\gamma\Delta\tilde{g}_A(W)\}\bigr| \leqslant \gamma\sigma^{-2} \leqslant \sigma^{-2}. \tag{4.6}$$

To obtain better estimates than in Poisson approximation, we proceed as Barbour and Čekanavičius (2002). To this end, let U and V be independent integer-valued random variables. Then it is easy to see that, for any bounded function F,

$$|\mathbb{E}\Delta F(U)| \leqslant 2\|F\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(U), \mathscr{L}(U+1)\bigr), \tag{4.7}$$

$$|\mathbb{E}\Delta^2 F(U+V)| \leqslant 4\|F\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(U), \mathscr{L}(U+1)\bigr)\, d_{\mathrm{TV}}\bigl(\mathscr{L}(V), \mathscr{L}(V+1)\bigr). \tag{4.8}$$

4.2. Proofs of the theorems.

Lemma 1. *Let Φ be a random variable with $\mathbb{E}\Phi = \mu$ and $\mathrm{Var}\,\Phi = \tau^2$, such that $\Phi_c = (\Phi-\mu)/\tau$ satisfies* (2.1) *for some $\varepsilon \geqslant 0$. Then, for any random variable Z obeying $\mathbb{E}(Z|\Phi) = 0$ and $\mathbb{E}(Z^2|\Phi) \leqslant 1$,*

$$\bigl|\mathbb{E}\bigl\{\tau^2 f'(\Phi+Z) - (\Phi-\mu)f(\Phi+Z)\bigr\}\bigr| \leqslant (\varepsilon\tau^3+\tau^2+\tfrac{1}{2}\tau)\|f''\|, \qquad \textit{for all } f \in \mathcal{F}. \tag{4.9}$$

Proof. We write (2.1) in the form

$$\left|\mathbb{E}\{\tau^2 f'(\Phi) - (\Phi-\mu)f(\Phi)\}\right| \leqslant \varepsilon\tau^3\|f''\|, \qquad \text{for all } f\in\mathcal{F}. \tag{4.10}$$

By Taylor expansion of f around Φ we obtain

$$\begin{aligned}\mathbb{E}\{\tau^2 f'(\Phi+Z) - (\Phi-\mu)f(\Phi+Z)\} &= \mathbb{E}\Big\{\tau^2\Big[f'(\Phi) + Z\int_0^1 f''(\Phi+sZ)\,ds\Big] \\ &\quad - (\Phi-\mu)\Big[f(\Phi)+Zf'(\Phi)+Z^2\int_0^1(1-s)f''(\Phi+sZ)\,ds\Big]\Big\}.\end{aligned}$$

With $\mathbb{E}\{(\Phi-\mu)Zf'(\Phi)\}=0$ the estimate is easily obtained. ☐

Proof of Theorem 1. First, we prove inequality (2.2). Let Z' be a random variable with $\mathscr{L}(Z'|\Phi) = \mathrm{Po}(\gamma'_\Phi)$, where $\gamma'_\Phi = \langle\Phi-\lambda\rangle$, and let $Y\sim\mathrm{Po}(\lambda)$ be independent of (Φ, Z'). Set $Z = Z'-\gamma'_\Phi$ and $W = \Phi+Z+(Y-\lambda)$. Then, $W\sim\mathrm{TP}[\mathscr{L}(\Phi)\times\delta_\lambda]$, and, with $s = \lfloor\mu-\tau^2-\lambda\rfloor$ and $\gamma = \langle\mu-\tau^2-\lambda\rangle$, taking $\sigma^2=\tau^2+\lambda$ in (4.5), we have

$$\begin{aligned}\mathbb{E}\mathcal{A}g_A(W-s) &= \mathbb{E}\{(\tau^2+\lambda)\Delta\tilde g_A(W) - (W-\mu)\tilde g_A(W) + \gamma\Delta\tilde g_A(W)\} \\ &= \mathbb{E}\{\tau^2\Delta\tilde g_A(W) - (\Phi-\mu)\tilde g_A(W)\} + \mathbb{E}\{(\gamma-\gamma'_\Phi)\Delta\tilde g_A(W)\} \\ &= \mathbb{E}\{\tau^2\Delta h_A(\Phi+Z-\lambda) - (\Phi-\mu)h_A(\Phi+Z-\lambda)\} \\ &\quad + \mathbb{E}\{(\gamma-\gamma'_\Phi)\Delta\tilde g_A(W)\},\end{aligned} \tag{4.11}$$

where for the second equality we use the fact that

$$\mathbb{E}\{Yg(Y)\} = \mathbb{E}\{\lambda g(Y+1)\} \tag{4.12}$$

(see Barbour et al. (1992, p. 5)) and for the third equality we put

$$h_A(j) := \mathbb{E}\{\tilde g_A(W) \mid \Phi+Z-\lambda = j\} = \mathbb{E}\{\tilde g_A(j+Y)\}$$

and use the independence of Y.

The second term in (4.11) is simply estimated with (4.3). To estimate the main term we use (4.9) for an appropriate interpolation function h_A.

Hence, we construct a function $f_A\in\mathcal{F}$, satisfying the conditions $f_A(j) = h_A(j)$ and $f'_A(j) = \Delta h_A(j)$ for all $j\in\mathbb{Z}$. For $j\in\mathbb{Z}$ and $x\in[0,1)$ define the function

$$\begin{aligned}&f_A(j+x) := \\ &\quad h_A(j)+\Delta h_A(j)x+\Delta^2 h_A(j)\cdot\begin{cases}-c_0x^2/2 & \text{if } x\leqslant c_0^{-1}2^{-1/2}\\ c_0(1-x)(3-2\sqrt{2}-x)/2 & \text{if } x> c_0^{-1}2^{-1/2},\end{cases}\end{aligned}$$

where $c_0 = 1+\sqrt{2}$. Clearly, f satisfies the desired conditions, and we can then use calculus to show that

$$\|f''_A\| \leqslant c_0\|\Delta^2 h_A\|.$$

The interpolation of h with the function f is optimal in the sense that the factor c_0 cannot be improved in the above inequality.

Using (4.7) for $F := \Delta\tilde{g}_A$ and invoking the bounds (4.3), we have

$$\begin{aligned}\bigl|\Delta^2 h_A(j)\bigr| &= \bigl|\mathbb{E}\bigl\{\Delta^2\tilde{g}_A(j+Y)\bigr\}\bigr| \leqslant 2\|\Delta\tilde{g}_A\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(Y),\mathscr{L}(Y+1)\bigr)\\ &\leqslant \frac{2}{(\tau^2+\lambda)\sqrt{\lambda}},\end{aligned}\tag{4.13}$$

where we have used the fact that $d_{\mathrm{TV}}\bigl(\mathscr{L}(Y),\mathscr{L}(Y+1)\bigr) \leqslant 1/\sqrt{\lambda}$, which can easily be proved with Stein's method for the Poisson case using (4.12).

Applying Lemma 1 to (4.11) with $f(x) = f_A(x-\lambda)$ we obtain the final bound

$$\begin{aligned}\bigl|\mathbb{E}\mathcal{A}g_A(W-s)\bigr| &\leqslant (\varepsilon\tau^3+\tau^2+\tfrac{1}{2}\tau)\|f_A''\| + \|\Delta\tilde{g}_A\|\\ &\leqslant \frac{c_0(2\varepsilon\tau^3+2\tau^2+\tau)}{(\tau^2+\lambda)\sqrt{\lambda}} + \frac{1}{\tau^2+\lambda}.\end{aligned}$$

As $\operatorname{Var} W \leqslant \tau^2+\lambda+1$, it follows from Chebyshev's inequality that

$$\mathbb{P}\bigl[W-s<0\bigr] \leqslant \left\{\frac{\operatorname{Var} W}{(\tau^2+\lambda)^2}\wedge 1\right\} \leqslant \left\{\left(\frac{1}{\tau^2+\lambda}+\frac{1}{(\tau^2+\lambda)^2}\right)\wedge 1\right\} \leqslant \frac{2}{\tau^2+\lambda},$$

and hence, from (4.1), inequality (2.2) is proved.

For inequality (2.3), write $Y = Y_1+Y_2$, where Y_1, Y_2 are independent, $\mathrm{Po}(\lambda/2)$ distributed random variables. Using (4.8) for $F := \tilde{g}_{\{k\}}$ and invoking (4.4), we replace the estimate (4.13) by

$$\bigl|\Delta^2 h_{\{k\}}(j)\bigr| \leqslant 4\|\tilde{g}_{\{k\}}\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(Y_1),\mathscr{L}(Y_1+1)\bigr)^2 \leqslant \frac{8}{(\tau^2+\lambda)\lambda}. \qquad \square$$

Proof of Theorem 2. We first prove (2.5). Write $X = (\Phi,\Lambda)$. Given X fixed, let $Y \sim \mathrm{Po}(\Lambda)$ and $Z' \sim \mathrm{Po}(\gamma')$ be independent, where $\gamma' = \langle\Phi-\Lambda\rangle$, and set $W = \Phi + (Z'-\gamma') + (Y-\Lambda)$. Then, $\mathscr{L}(W|X) = \mathrm{TP}(\Phi,\Lambda)$; we now use (4.1) with the conditional distribution $\mathbb{P}^X$ of W given X with $\mu = \Phi$ and $\sigma^2 = \lambda = \mathbb{E}\Lambda$ to obtain our estimate. From (4.5), with $s = \lfloor\Phi-\lambda\rfloor$ and $\gamma = \langle\Phi-\lambda\rangle$, it follows that

$$\begin{aligned}\mathbb{E}^X\mathcal{A}g_A(W-s) &= \mathbb{E}^X\bigl\{\lambda\Delta\tilde{g}_A(W) - (W-\Phi)\tilde{g}_A(W)\bigr\} + \gamma\mathbb{E}^X\Delta\tilde{g}_A(W)\\ &= \mathbb{E}^X\bigl\{(\lambda-\Lambda)\Delta\tilde{g}_A(W)\bigr\} + \mathbb{E}^X\bigl\{(\gamma-\gamma')\Delta\tilde{g}_A(W)\bigr\},\end{aligned}$$

where we have used (4.12) for $Y+Z' \sim \mathrm{Po}(\Lambda+\gamma')$ and hence, using (4.3),

$$\bigl|\mathbb{E}^X\mathcal{A}g_A(W-s)\bigr| \leqslant \lambda^{-1}(|\lambda-\Lambda|+1).$$

Moreover, by Chebyshev's inequality,

$$\mathbb{P}^X\bigl[W-s<0\bigr] \leqslant \frac{\lambda+\gamma}{\lambda^2}.\tag{4.14}$$

Hence, we can bound (4.1) to give

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W|X),\mathrm{TP}(\Phi,\Lambda)\bigr) \leqslant \frac{|\lambda-\Lambda|}{\lambda} + \frac{1}{\lambda} + \frac{\lambda+\gamma}{\lambda^2}.$$

Taking expectation over X, the claim follows.

To prove inequality (2.6), use (4.7) for $F := \tilde{g}_{\{k\}}$ and the bound (4.4) to obtain

$$\begin{aligned}\bigl|\mathbb{E}^X\bigl\{\mathcal{A}g_{\{k\}}(W-s)\bigr\}\bigr| &\leqslant 2\bigl(|\lambda-\Lambda|+1\bigr)\|\tilde{g}_{\{k\}}\|\, d_{\mathrm{TV}}\bigl(\mathscr{L}(Y),\mathscr{L}(Y+1)\bigr)\\ &\leqslant \frac{2\bigl(|\lambda-\Lambda|+1\bigr)}{\lambda\sqrt{\Lambda}}.\end{aligned}$$

By Chebyshev's inequality, we obtain

$$\begin{aligned}\mathbb{E}\biggl\{1\wedge\frac{2\bigl(|\lambda-\Lambda|+1\bigr)}{\lambda\sqrt{\Lambda}}\biggr\} &\leqslant \mathbb{E}\biggl\{I[\Lambda\leqslant\lambda/2]+I[\Lambda>\lambda/2]\frac{2\sqrt{2}(|\lambda-\Lambda|+1)}{\lambda^{3/2}}\biggr\}\\ &\leqslant \frac{4\nu^2}{\lambda^2}+\frac{2\sqrt{2}(\nu+1)}{\lambda^{3/2}}\end{aligned}$$

and hence, with (4.14) and (4.1), the claim. □

Acknowledgements

I thank Andrew Barbour and Małgorzata Roos for helpful discussions and the referees for their comments which improved some of the results and the presentation.

References

Barbour, A.D. and V. Čekanavičius. 2002. *Total variation asymptotics for sums of independent integer random variables*, Ann. Probab. **30**, 509–545.

Barbour, A.D., L. Holst, and S. Janson. 1992. *Poisson approximation*, Oxford University Press.

Barbour, A.D. and S. Utev. 2004. *Approximating the Reed–Frost epidemic process*, Stochastic Process. Appl. **113**, 173–197.

Barbour, A.D. and A. Xia. 1999. *Poisson perturbations*, ESAIM Probab. Statist. **3**, 131–159.

Burgess, D. and D. McDonald. 1995. *An elementary proof of the local central limit theorem*, J. Theoret. Probab. **8**, 693–701.

Čekanavičius, V. and P. Vaĭtkus. 2001. *Centered Poisson approximation by the Stein method*, Lithuanian Math. J. **41**, 319–329.

McDonald, D. 1979. *On local limit theorem for integer valued random variables*, Theory Probab. Appl. **24**, 613–619.

Reinert, G. 1998. *Coupling for normal approximations with Stein's method*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **41**, 193–207.

Stein, C. 1972. *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, pp. 583–602.

submitted to the Annals of Probability

SYMMETRIC AND CENTERED BINOMIAL APPROXIMATION OF SUMS OF LOCALLY DEPENDENT RANDOM VARIABLES

By Adrian Röllin*

Stein's method is used to approximate sums of discrete and locally dependent random variables by a centered and symmetric Binomial distribution. Under appropriate smoothness properties of the summands, the same order of accuracy as in the Berry-Essen Theorem is achieved. The approximation of the total number of points of a point processes is also considered. The results are applied to the exceedances of the r-scans process and to the Matérn hardcore point process type I.

1. Introduction

The approximation of sums of dependent random variables by the standard normal distribution has been investigated in a large variety of settings. The accuracy of approximation is most often measured by the Kolmogorov and Wasserstein metrics. The use of stronger metrics typically requires that some 'smoothness'-condition must be satisfied.

In this paper, under the assumption of a general local dependence structure, we study the approximation of sums of discrete random variables by a symmetric and centered Binomial distribution. This distribution serves as replacement for the normal distribution in a discrete setting. Under some general smoothness property of the summands, the same order of accuracy as in the Berry-Essen Theorem can be achieved, but now for the total variation metric. We also examine another metric, from which local limit approximations can be obtained.

In the setting of independent summands, approximation by a centered Poisson distribution has been successfully adopted by Čekanavičius and Vaĭtkus (2001) and Barbour and Čekanavičius (2002). However, for dependent summands, applications were limited to simple examples; first attempts were made by Barbour and Xia (1999) and Čekanavičius and Vaĭtkus (2001). In contrast, the results in this paper are of general nature and allow a wide range of applications.

The proofs are based on Stein's method for distributional approximation. A main idea, introduced in Röllin (2005), is to use interpolation functions to represent the Stein operator of a discrete distribution as the Stein operator of a continuous distribution. In the case of the Binomial, this then allows the application of standard techniques in Stein's method for normal approximation. A careful analysis of the remainder terms then shows how a suitable smoothness condition can be exploited, to obtain total variation error bounds.

The paper is organized as follows. In the next section, we introduce the main technique in the simple case of independent summands. In section 3 these results

*Partially supported by Schweizerischer Nationalfondsprojekt 20-107935/1

AMS 2000 subject classifications: Primary 60F05

Keywords and phrases: Stein's method; total variation metric; Binomial distribution; local dependence

are extended to locally dependent summands and section 4 shows their application in some examples. Section 5 contains some technical lemmas.

1.1. Notation. Denote by $\mathrm{Bi}(n,p)$ the Binomial distribution with n trials of probability p each. Denote by $\widehat{\mathrm{Bi}}(n,p)$ the centered Binomial distribution, i.e. a Binomial distribution shifted by $-np$. Note that this distribution does not necessarily lie on the integers, but on a lattice of $\mathbb{R}$ with span 1.

Throughout the paper, we shall be concerned with two metrics for probability distributions, the total variation metric d_{TV} and the local limit metric d_{loc}, where, for two probability distributions P and Q,

$$\begin{aligned} d_{\mathrm{TV}}\bigl(P,Q\bigr) &:= \sup_{A\subset\mathbb{R}}\bigl|P(A)-Q(A)\bigr|,\\ d_{\mathrm{loc}}\bigl(P,Q\bigr) &:= \sup_{x\in\mathbb{R}}\bigl|P\bigl([x,x+1)\bigr)-Q\bigl([x,x+1)\bigr)\bigr|. \end{aligned}$$

For simplicity, we will often use the notation d_l, where $l=1$ will stand for d_{TV} and $l=2$ for d_{loc}.

We denote by $\|\cdot\|$ the supremum norm if applied to functions, and the variation norm if applied to measures. Let δ_x denote the unit mass at $x\in\mathbb{R}$, and $*$ the convolution of measures. Define for any measure μ and any $l\in\mathbb{N}:=\{1,2,\dots\}$

$$D^l(\mu)=\bigl\|\mu*(\delta_1-\delta_0)^{*l}\bigr\|.$$

Note that for measures μ and λ,

$$D^1\bigl(\mu\bigr)=2\,d_{\mathrm{TV}}\bigl(\mu,\mu*\delta_1\bigr), \tag{1.1}$$

$$D^2(\mu*\lambda)\leqslant D^1(\mu)D^1(\lambda). \tag{1.2}$$

Furthermore, define $\langle x\rangle := x-\lfloor x\rfloor$ to be the fractional part of $x\in\mathbb{R}$, and $(x)_+ = x\vee 0$.

1.2. Basic setup. Consider a sum of the form $W=\sum_{i\in J}\xi_i$, where W takes its values in a lattice of $\mathbb{R}$ with span 1. The expectation of W has no influence on the quality of the approximation, and we therefore assume without loss of generality that $\mathbb{E}W=0$; this can always be accomplished by subtracting the expectation from each individual summand. Each of the summands may now take its values on a different lattice; this, however, will result in no further complications.

To approximate W by a centered binomial distribution, we have to choose n in such a way that the variance of $\widehat{\mathrm{Bi}}(n,1/2)$ is as close to the variance of W as possible. As n has to be integer, this is only possible up to a rounding error. However, the symmetric and centered Binomial distribution thus chosen will in general take its values on a different lattice from W and the total variation distance will become 1. To circumvent this problem, we introduce an additional parameter t and approximate W by a centered Binomial distribution with success probability $1/2-t$ instead (t being small), to be able to match not only the variance but also the lattice.

Hence, to put the above in a rigorous form, we will make the following assumptions if not otherwise stated:

Assumptions G: Let J be a finite set and let $\{\xi_i, i \in J\}$ be a collection of random variables with $\mathbb{E}\xi_i = 0$ for all $i \in J$ and assume that there are numbers $\{a_i \in \mathbb{R}; i \in J\}$ such that almost surely $\xi_i \in \mathbb{Z} + a_i$. Let $W = \sum_{i\in J}\xi_i$; then $\mathbb{E}W = 0$ and almost surely $W \in \mathbb{Z} + a$ for $a := \sum_{i\in J} a_i$. Assume that $\sigma^2 := \operatorname{Var} W > 1$. Define now $\delta := \langle -4\sigma^2 \rangle$ and $t := \langle a + 2\sigma^2 + \delta/2 \rangle/(4\sigma^2 + \delta)$. Clearly, $4\sigma^2 + \delta = \lceil 4\sigma^2 \rceil$, and by definition the distribution $\widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2 \rceil, 1/2 - t\big)$ has expectation 0; it is also easy to check that it takes values in $\mathbb{Z} + a$.

From the above definition, we see that t is only of order $\mathrm{O}(\sigma^{-2})$, which is rather small in the setting that we are concerned with; Corollary 2.3 shows how to obtain results without t, using Lemma 5.2.

2. Sum of Independent Random Variables

First, we examine the case of independent discrete summands. Previous work on total variation approximation has been concerned with the compound Poisson distribution (see Le Cam (1965) and Roos (2003) and references therein), the signed compound Poisson distribution (see Čekanavičius (1997) and references therein), the Poisson distribution (see Barbour et al. (1992)), the centered Poisson distribution (see Čekanavičius (1998), Čekanavičius and Vaĭtkus (2001), Barbour and Xia (1999) and Barbour and Čekanavičius (2002)) and some more general distributions (see Brown and Xia (2001)).

We present the theorem below to demonstrate the main technique in a simple setting, noting that it also follows as a consequence of Theorem 3.1.

Theorem 2.1. *Let $\{\xi_i; i \in J\}$ be independent and satisfy Assumptions G. Then, if the ξ_i have finite third moments,*

$$d_l\big(\mathscr{L}(W), \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2\rceil, 1/2 - t\big)\big) \leqslant \sigma^{-2}\Big(\sum_{i\in J} c_{l,i}\rho_i + 1.75\Big), \qquad l = 1, 2,$$

where $\rho_i = \sigma_i^3 + \frac{1}{2}\mathbb{E}|\xi_i|^3$, $\sigma_i^2 = \operatorname{Var}\xi_i$ and $c_{l,i} = D^l\big(\mathscr{L}(W - \xi_i)\big)$.

It is clear that the above bound is useful only if the $c_{l,i}$ are small. In the case of n identically distributed random variables, we need $c_{1,i} = \mathrm{o}(1)$ as $n \to \infty$ for asymptotic approximation in total variation, and in order to deduce a local limit theorem we must have $c_{2,i} = \mathrm{o}(n^{-1/2})$. This is however always the case if $D^1(X_1) < 2$ (this corresponds to the usual condition in the LLT that X_1 must not be concentrated on a lattice with span greater than 1), as can be seen from (5.9)–(5.10), and we then even have $c_{l,i} = \mathrm{O}(n^{-l/2})$ for $l = 1, 2$.

Before proving the theorem, we start with a short summary of Stein's method for Binomial approximation; for details see also Stein (1986) and Ehm (1991). Denote by $F(M)$ the set of all real valued measurable functions on some given measure space M. A Stein operator $\mathcal{B} : F(\mathbb{Z}) \to F(\mathbb{Z})$ for the Binomial distribution $\mathrm{Bi}(n,p)$ is characterized by the fact that, for any integer valued random variable W,

$$\mathbb{E}(\mathcal{B}g)(W) = 0 \text{ for all bounded } g \in F(\mathbb{Z}) \iff W \sim \mathrm{Bi}(n,p), \tag{2.1}$$

and a possible choice is

$$(\mathcal{B}g)(z) = qzg(z-1) - p(n-z)g(z), \qquad \text{for all } z \in \mathbb{Z}, \tag{2.2}$$

where, as usual, we put $q = 1-p$.

Let $h \in F(\mathbb{Z})$ be a bounded function. Then, the solution $g = g_h$ to the Stein equation

$$(\mathcal{B}g)(z) = I[0 \leqslant z \leqslant n]\bigl\{h(z) - \mathbb{E}h(Y)\bigr\}, \qquad \text{for all } z \in \mathbb{Z}, \tag{2.3}$$

where $Y \sim \mathrm{Bi}(n,p)$, is also bounded. If the functions h are of the form $h(z) = h_A(z) = I[z \in A]$, $A \subset \mathbb{Z}$, we have the uniform bound

$$\|\Delta g_A\| \leqslant \frac{1 - p^{n+1} - q^{n+1}}{(n+1)pq}, \tag{2.4}$$

where $\Delta g(z) := g(z+1) - g(z)$, and the same bound holds for $\|g_{\{b\}}\|$, $b \in \mathbb{Z}$; see Ehm (1991). Now, for all $z \in \mathbb{Z}$, we can write

$$I[z \in A] - \mathbb{P}[Y \in A] = (\mathcal{B}g_A)(z) + I[z \notin \{0 \dots n\}]\bigl(I[z \in A] - \mathbb{P}[Y \in A]\bigr),$$

and thus, for any integer valued random variable V,

$$\begin{aligned} d_{\mathrm{TV}}\bigl(\mathscr{L}(V), \mathrm{Bi}(n,p)\bigr) &= \sup_{A \subset \mathbb{Z}} \bigl|\mathbb{P}[V \in A] - \mathbb{P}[Y \in A]\bigr| \\ &\leqslant \sup_{A \subset \mathbb{Z}} \bigl|\mathbb{E}(\mathcal{B}g_A)(V)\bigr| + \mathbb{P}\bigl[|V - n/2| > n/2\bigr]. \end{aligned} \tag{2.5}$$

We now construct a Stein operator for the centered Binomial distribution $\widehat{\mathrm{Bi}}(n,p)$ on the lattice $\mathbb{Z} - np$. For any function $g \in F(\mathbb{Z})$ define the function $\hat{g} \in F(\mathbb{Z} - np)$ by $\hat{g}(w) := g(w + np)$ for $w \in \mathbb{Z} - np$. Then the Stein operator is defined as

$$\begin{aligned} (\hat{\mathcal{B}}\hat{g})(w) &:= (\mathcal{B}g)(w + np) \\ &= p(w+np)g(w+np) + q(w+np)g(w-1+np) - npg(w+np) \\ &= w\bigl(p\hat{g}(w) + q\hat{g}(w-1)\bigr) - npq\Delta\hat{g}(w-1). \end{aligned} \tag{2.6}$$

for all $w \in \mathbb{Z} - np$. Thus, for $W = V - np$, an inequality corresponding to (2.5) holds, namely

$$\begin{aligned} &d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \widehat{\mathrm{Bi}}(n,p)\bigr) \\ &\qquad \leqslant \sup_{B \subset \mathbb{Z} - np} \bigl|\mathbb{E}(\hat{\mathcal{B}}\hat{g}_B)(W)\bigr| + \mathbb{P}\bigl[|W + n(p - 1/2)| > n/2\bigr]. \end{aligned} \tag{2.7}$$

An equivalent inequality holds for the d_{loc} metric, but the supremum is taken only over the sets $\{b\}$, $b \in \mathbb{Z} - np$.

Under the assumptions of the theorem, $n = \lceil 4\sigma^2 \rceil = 4\sigma^2 + \delta$ and $p = 1/2 - t$, and (2.6) becomes

$$(\hat{\mathcal{B}}\hat{g})(w) = w\Theta\hat{g}(w-1) - \sigma^2\Delta\hat{g}(w-1) + \bigl(t^2(4\sigma^2+\delta) - wt - \delta/4\bigr)\Delta\hat{g}(w-1), \quad (2.8)$$

where $\Theta\hat{g}(w) := \frac{1}{2}\bigl(\hat{g}(w+1) + \hat{g}(w)\bigr)$. Since $\sigma^2 > 1$, the bound (2.4) simplifies to

$$\|\Delta\hat{g}_B\| \leqslant \frac{1}{\sigma^2}. \quad (2.9)$$

To see this, note that $t < 1/\lceil 4\sigma^2 \rceil = 1/n$ and $n = \lceil 4\sigma^2 \rceil \geqslant 5$. Then from (2.4) we have

$$\|\Delta\hat{g}_B\| \leqslant \frac{1}{(n+1)pq} = \frac{1}{(n+1)(1/4-t^2)} \leqslant \frac{4n^2}{(n+1)(n^2-4)} \leqslant \frac{4}{n} \leqslant \frac{1}{\sigma^2}.$$

Lemma 2.2. *Assume the conditions of Theorem 2.1. Define* $\mathcal{A} : F(\mathbb{Z}+a) \to F(\mathbb{Z}+a)$ *by*

$$(\mathcal{A}\hat{g})(w) := w\Theta\hat{g}(w-1) - \sigma^2\Delta\hat{g}(w-1), \qquad w \in \mathbb{Z}+a,\ \hat{g} \in F(\mathbb{Z}+a).$$

Then,

$$\bigl|\mathbb{E}(\mathcal{A}\hat{g})(W)\bigr| \leqslant \Bigl(\|\Delta\hat{g}\| \sum_{i\in J} c_{1,i}\rho_i\Bigr) \wedge \Bigl(\|\hat{g}\| \sum_{i\in J} c_{2,i}\rho_i\Bigr). \quad (2.10)$$

Proof. For every $w \in \mathbb{Z}+a$ and $x \in [0,1)$ define

$$f(w+x) := \Theta\hat{g}(w-1) + x\Delta\hat{g}(w-1) + \tfrac{1}{2}x^2\Delta^2\hat{g}(w-1). \quad (2.11)$$

One easily checks that $f \in C^1$ and $f(w) = \Theta\hat{g}(w-1)$ and $f'(w) = \Delta\hat{g}(w-1)$, hence

$$(\mathcal{A}\hat{g})(w) = wf(w) - \sigma^2 f'(w), \quad (2.12)$$

for all $w \in \mathbb{Z}+a$. Furthermore, f' is absolutely continuous, hence f'' exists almost everywhere. Choose f'' to be the function

$$f''(w+x) = \Delta^2\hat{g}(w-1) \quad (2.13)$$

for all $w \in \mathbb{Z}+a$, $0 \leqslant x < 1$.

We can now apply the usual Taylor expansion (cf. Reinert (1998), Theorem 2.1), but with a refined estimate of the remainder terms. Write $W_i = W - \xi_i$, $i \in J$; then

$$\xi_i f(W) = \xi_i f(W_i) + \xi_i^2 f'(W_i) + \xi_i^3 \int_0^1 (1-s) f''(W_i + s\xi_i)\, ds,$$

$$\sigma_i^2 f'(W) = \sigma_i^2 f'(W_i) + \xi_i\sigma_i^2 \int_0^1 f''(W_i + s\xi_i)\, ds,$$

and hence, using the independence of ξ_i and W_i and that $\mathbb{E}\xi_i = 0$,

$$\begin{aligned}\bigl|\mathbb{E}\bigl\{\xi_i f(W) - \sigma_i^2 f'(W)\bigr\}\bigr| \leqslant \mathbb{E}\biggl|\xi_i^3 \int_0^1 &(1-s)\mathbb{E}\bigl[f''(W_i+s\xi_i) \bigm| \xi_i\bigr]\, ds \\ &-\xi_i\sigma_i^2 \int_0^1 \mathbb{E}\bigl[f''(W_i+s\xi_i) \bigm| \xi_i\bigr]\, ds\biggr|. \end{aligned} \quad (2.14)$$

Note now that for any real valued random variable U taking values on a lattice with span 1, we obtain together with (2.13)

$$\bigl|\mathbb{E}\bigl(f''(U+z)\bigr)\bigr| \leqslant \Bigl(\|\Delta\hat{g}\|D^1\bigl(\mathscr{L}(U)\bigr)\Bigr) \wedge \Bigl(\|\hat{g}\|D^2\bigl(\mathscr{L}(U)\bigr)\Bigr), \tag{2.15}$$

for all $z \in \mathbb{R}$. Thus, from (2.14) and (2.15),

$$\begin{aligned} &\bigl|\mathbb{E}\bigl\{\xi_i f(W) - \sigma_i^2 f'(W)\bigr\}\bigr| \\ &\leqslant \Bigl(\|\Delta\hat{g}\|D^1\bigl(\mathscr{L}(W_i)\bigr)\bigl(\sigma_i^3 + \tfrac{1}{2}\mathbb{E}|\xi_i|^3\bigr)\Bigr) \wedge \Bigl(\|\hat{g}\|D^2\bigl(\mathscr{L}(W_i)\bigr)\bigl(\sigma_i^3 + \tfrac{1}{2}\mathbb{E}|\xi_i|^3\bigr)\Bigr). \end{aligned} \tag{2.16}$$

Now, using (2.12) we have

$$\bigl|\mathbb{E}\bigl\{\mathcal{A}\hat{g}(W)\bigr\}\bigr| \leqslant \sum_{i\in J}\bigl|\mathbb{E}\bigl\{\xi_i f(W) - \sigma_i^2 f'(W)\bigr\}\bigr|$$

and with (2.16) the lemma is proved. □

Proof of Theorem 2.1. Recall that, by Assumptions G, the distributions $\mathscr{L}(W)$ and $\widehat{\mathrm{Bi}}\bigl(\lceil 4\sigma^2\rceil, 1/2 - t\bigr)$ are concentrated on the same lattice. Thus, using (2.7) and the form (2.8) of the Stein operator, and applying the left side of the minimum in (2.10) to the first part of (2.8) with the bound (2.9) gives

$$\begin{aligned} &d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \widehat{\mathrm{Bi}}(4\sigma^2+\delta, 1/2-t)\bigr) \\ &\quad\leqslant \frac{\sum_{i\in J} c_{1,i}\rho_i}{\sigma^2} + \frac{t^2(4\sigma^2+\delta) + \sigma t + \delta/4}{\sigma^2} + \mathbb{P}\bigl[|W| \geqslant 2\sigma^2 - 1\bigr]. \end{aligned} \tag{2.17}$$

To bound the middle part of (2.17) note that $0 \leqslant t < (4\sigma^2+\delta)^{-1}$ and $0 \leqslant \delta < 1$. Thus, recalling that $\sigma^2 > 1$, we obtain the simple bounds

$$t^2(4\sigma^2+\delta) < (4\sigma^2+\delta)^{-1} \leqslant 1/4, \qquad \sigma t \leqslant \sigma/(4\sigma^2+\delta) \leqslant 1/4, \qquad \delta/4 \leqslant 1/4.$$

Applying Chebyshev's inequality on the last term of (2.17) we obtain

$$\mathbb{P}\bigl[|W| \geqslant 2\sigma^2 - 1\bigr] \leqslant \frac{\sigma^2}{(2\sigma^2-1)^2} \leqslant \frac{1}{\sigma^2}.$$

The d_{loc} case is analogous, using the right side of the minimum in (2.10) instead and the remark after (2.4). □

Note that in the next corollary we do not assume that the ξ_i have expectation zero.

Corollary 2.3. *Let W be the sum of independent and integer valued random variables $\{\xi_i, i \in J\}$ with $\sigma_i^2 = \operatorname{Var}\xi_i$ and*

$$v_i = \min\bigl\{1/2, 1 - d_{\mathrm{TV}}\bigl(\mathscr{L}(\xi_i), \mathscr{L}(\xi_i+1)\bigr)\bigr\}.$$

Then, if $\sigma^2 > 1$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2) * \delta_s\big) \leqslant \frac{2\sum\big(\sigma_i^3 + \frac{1}{2}\mathbb{E}|\xi_i|^3\big)}{\sigma^2(V - v^*)^{1/2}} + \frac{1 + 2.25\sigma^{-1} + 0.25\sigma^{-2}}{\sigma},$$

$$d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2) * \delta_s\big) \leqslant \frac{8\sum\big(\sigma_i^3 + \frac{1}{2}\mathbb{E}|\xi_i|^3\big)}{\sigma^2(V - 4v^*)_+} + \frac{3.25 + 0.25\sigma^{-1}}{\sigma^2},$$

where $s := \lceil \mu - \lceil 4\sigma^2\rceil/2\rceil$, $\mu = \mathbb{E}W$, $V = \sum_{i\in J} v_i$ *and* $v^* = \max_{i\in J} v_i$.

Proof. Define $W_0 = W - \mu$, and let t be defined with respect to W_0, taking $a = -\mu$. Then, as the metrics d_l are shift invariant,

$$\begin{aligned} d_l\big(\mathscr{L}(W), \mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2) * \delta_s\big) &= d_l\big(\mathscr{L}(W_0), \mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2) * \delta_{s-\mu}\big) \\ &\leqslant d_l\big(\mathscr{L}(W_0), \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2\rceil, 1/2 - t\big)\big) + d_l\big(\mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2 - t), \mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2)\big) \\ &=: R_1^l + R_2^l, \end{aligned}$$

since $\mathrm{Bi}(\lceil 4\sigma^2\rceil, 1/2 - t) * \delta_s * \delta_{-\mu} = \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2\rceil, 1/2 - t\big)$.

Applying Lemma 5.2 to R_2^l with the fact that $0 \leqslant t \leqslant (4\sigma^2 + \delta)^{-1}$ gives

$$R_2^1 \leqslant \sigma^{-1}\big(1 + (2\sigma)^{-1} + (4\sigma^2)^{-1}\big), \qquad R_2^2 \leqslant \sigma^{-2}\big(1.5 + (4\sigma)^{-1}\big). \tag{2.18}$$

Define now $c_l = \max_{i\in J}\big\{D^l\big(\mathscr{L}(W - \xi_i)\big)\big\}$. Application of (5.9)-(5.10) yields

$$c_1 \leqslant \frac{2}{(V - v^*)^{1/2}}, \qquad c_2 \leqslant \frac{8}{(V - 4v^*)_+}. \tag{2.19}$$

Thus, application of Theorem 2.1 to R_1^l proves the corollary. □

3. Locally dependent random variables

In this section we present the main results of the paper. We exploit a finite local dependence structure as presented in Chen and Shao (2004). In the context of Stein's method for normal approximation, it has been successfully applied to a variety of problems; see for example Barbour et al. (1989), Dembo and Rinott (1996) and Barbour and Xia (2001). Note that Barbour et al. (1989) use a slightly more general dependence structure, often yielding crucial improvements when approximating sums of dissociated random variables by the normal distribution. The generalization of Theorem 3.1 is straightforward, yet somewhat tedious, and we therefore use the simpler dependence structure of Chen and Shao (2004); see the Appendix for the more general version, but without proof.

Let $\{\xi_i; i \in J\}$ be a collection of random variables satisfying Assumptions G. For convenience, let ξ_A denote $\{\xi_i; i \in A\}$ for every subset $A \subset J$. Assume further the following dependence structure: For every $i \in J$ there are subsets $A_i \subset B_i \subset J$ such that ξ_i is independent of $\xi_{A_i^c}$, and ξ_{A_i} is independent of $\xi_{B_i^c}$. Define $\eta_i = \sum_{j\in A_i} \xi_j$ and $\tau_i = \sum_{j\in B_i} \xi_j$.

Theorem 3.1. *With W as above,*

$$d_l\big(\mathscr{L}(W), \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2\rceil, 1/2 - t\big)\big) \leqslant \sigma^{-2}\bigg(\sum_{i\in J}\vartheta_{l,i} + 1.75\bigg), \qquad l = 1, 2, \tag{3.1}$$

where

$$\begin{aligned}\vartheta_{l,i} = {}& \tfrac{1}{2}\mathbb{E}\big\{|\xi_i|\eta_i^2 D^l\big(\mathscr{L}(W|\xi_i,\eta_i)\big)\big\} + \mathbb{E}\big\{|\xi_i\eta_i(\tau_i - \eta_i)|D^l\big(\mathscr{L}(W|\xi_i,\eta_i,\tau_i)\big)\big\}\\ & + |\mathbb{E}\xi_i\eta_i|\mathbb{E}\big\{|\tau_i|D^l\big(\mathscr{L}(W|\tau_i)\big)\big\}\end{aligned} \tag{3.2}$$

If further there are constants $c_{l,i}$ such that almost surely

$$D^l\big(\mathscr{L}(W|\xi_{B_i})\big) \leqslant c_{l,i}, \tag{3.3}$$

then

$$\vartheta_{l,i} \leqslant c_{l,i}\big(\tfrac{1}{2}\mathbb{E}|\xi_i\eta_i^2| + \mathbb{E}|\xi_i\eta_i(\tau_i - \eta_i)| + |\mathbb{E}\xi_i\eta_i|\mathbb{E}|\tau_i|\big). \tag{3.4}$$

Proof. Estimate (3.4) is immediate. Following the proof of Theorem 2.1 and using Lemma 3.2 below, (3.1) is proved. □

Note that Theorem 2.1 follows from Theorem 3.1 with the choices $A_i = B_i = \{i\}$.

Lemma 3.2. *Assume the conditions of Theorem 3.1. Define $\mathcal{A} : F(\mathbb{Z} + a) \to F(\mathbb{Z} + a)$ as in Lemma 2.2. Then,*

$$\big|\mathbb{E}(\mathcal{A}\hat{g})(W)\big| \leqslant \Big(\|\Delta\hat{g}\|\sum_{i\in J}\vartheta_{1,i}\Big) \wedge \Big(\|\hat{g}\|\sum_{i\in J}\vartheta_{2,i}\Big). \tag{3.5}$$

Proof. We follow the proof of Lemma 2.2 right up to the end of the paragraph of (2.13). Note now that

$$\sigma^2 = \sum_{i\in J}\mathbb{E}\{\xi_i\eta_i\} \tag{3.6}$$

and that, by Taylor expansion, almost surely

$$\begin{aligned}\xi_i f(W) &= \xi_i f(W - \eta_i) + \xi_i\eta_i f'(W - \eta_i) + \xi_i\eta_i^2\int_0^1 f''(W - \eta_i + s\eta_i)\,ds,\\ \xi_i\eta_i f'(W - \eta_i) &= \xi_i\eta_i f'(W - \tau_i) + \xi_i\eta_i(\tau_i - \eta_i)\int_0^1 f''(W - \eta_i + s(\tau_i - \eta_i))\,ds,\\ \mathbb{E}\{\xi_i\eta_i\}f'(W) &= \mathbb{E}\{\xi_i\eta_i\}f'(W - \tau_i) + \mathbb{E}\{\xi_i\eta_i\}\tau_i\int_0^1 f''(W + s\tau_i)\,ds.\end{aligned} \tag{3.7}$$

Now, using the facts that $\mathbb{E}\xi_i = 0$, that ξ_i is independent of $W - \eta_i$ and that η_i is independent of $W-\tau_i$, we obtain from (3.6) and (3.7) that

$$\begin{aligned}
\mathbb{E}\bigl\{Wf(W)-\sigma^2f'(W)\bigr\} &= \sum_{i\in J}\mathbb{E}\bigl\{\xi_i f(W)-\mathbb{E}\{\xi_i\eta_i\}f'(W)\bigr\}\\
= \sum_{i\in J}\mathbb{E}\biggl\{\xi_i\eta_i^2\int_0^1 &(1-s)\mathbb{E}\bigl\{f''(W-\eta_i+s\eta_i)\bigm|\xi_i,\eta_i\bigr\}\,ds\\
&+\xi_i\eta_i(\tau_i-\eta_i)\int_0^1\mathbb{E}\bigl\{f''(W-\tau_i+s(\tau_i-\eta_i))\bigm|\xi_i,\eta_i,\tau_i\bigr\}\,ds\\
&-\mathbb{E}\bigl\{\xi_i\eta_i\bigr\}\tau_i\int_0^1\mathbb{E}\bigl\{f''(W-\tau_i+s\tau_i))\bigm|\tau_i\bigr\}\,ds\biggr\}.
\end{aligned}$$

With (2.12) and (2.15) the lemma follows. □

We now give a point process version of Theorem 3.1, exploiting mainly the same dependency structure as before.

Theorem 3.3. *Let Φ be a simple point process on a Polish space J with mean measure μ. For all points $\alpha\in J$, assume that there are measurable subsets $A_\alpha\subset B_\alpha\subset J$, such that for every $\alpha\in J$*

$$\mathscr{L}\bigl(\Phi_\alpha(A_\alpha^c)\bigr)=\mathscr{L}\bigl(\Phi(A_\alpha^c)\bigr), \tag{3.8}$$

$$\Phi_\alpha(A_\alpha)\ \textit{and}\ \Phi_\alpha(B_\alpha^c)\ \textit{are independent}, \tag{3.9}$$

$$\Phi(A_\alpha)\ \textit{and}\ \Phi(B_\alpha^c)\ \textit{are independent}, \tag{3.10}$$

where Φ_α denotes the Palm process at point α. Then, for $W=\Phi(J)-\mu(J)$ and if $\sigma^2>1$,

$$\begin{aligned}
&d_l\bigl(\mathscr{L}(W),\widehat{\mathrm{Bi}}\bigl(\lceil 4\sigma^2\rceil,1/2-t\bigr)\bigr)\\
&\qquad\leqslant\sigma^{-2}\int_{\alpha\in J}\vartheta_l(\alpha)\mu(d\alpha)+1.75\sigma^{-2},\qquad l=1,2.
\end{aligned}\tag{3.11}$$

where, with $\Phi':=\Phi-\mu$ and $\Phi'_\alpha:=\Phi_\alpha-\mu$,

$$\begin{aligned}
\vartheta_l(\alpha)=\bigl|\mathbb{E}\Phi'_\alpha(A_\alpha)\bigr|&\mathbb{E}\bigl\{|\Phi'(B_\alpha)|D^l\bigl(\Phi(B_\alpha^c)\bigm|\Phi(B_\alpha)\bigr)\bigr\}\\
&+\tfrac12\mathbb{E}\bigl\{\Phi'_\alpha(A_\alpha)^2D^l\bigl(\Phi_\alpha(A_\alpha^c)\bigm|\Phi_\alpha(A_\alpha)\bigr)\bigr\}\\
&+\tfrac12\mathbb{E}\bigl\{\Phi'(A_\alpha)^2D^l\bigl(\Phi(A_\alpha^c)\bigm|\Phi(A_\alpha)\bigr)\bigr\}\\
&+\mathbb{E}\bigl\{|\Phi'_\alpha(A_\alpha)\Phi'_\alpha(B_\alpha\setminus A_\alpha)|D^l\bigl(\Phi_\alpha(B_\alpha^c)\bigm|\Phi_\alpha(A_\alpha),\Phi_\alpha(B_\alpha)\bigr)\bigr\}\\
&+\mathbb{E}\bigl\{|\Phi'(A_\alpha)\Phi'(B_\alpha\setminus A_\alpha)|D^l\bigl(\Phi(B_\alpha^c)\bigm|\Phi(A_\alpha),\Phi(B_\alpha)\bigr)\bigr\}.
\end{aligned}\tag{3.12}$$

Proof. Following the proof of Theorem 2.1 and Lemma 2.2, it is clear that we only have to bound $\mathbb{E}\bigl\{Wf(W)-\sigma^2f'(W)\bigr\}$ for f defined as in (2.11). In what follows, all integrals are taken over $\{\alpha\in J\}$ if not otherwise stated. Note first that, because of (3.8),

$$\sigma^2=\mathbb{E}\{\Phi(J)\Phi'(J)\}=\int\mu(d\alpha)\mathbb{E}\{\Phi'_\alpha(A_\alpha)+\Phi'_\alpha(A_\alpha^c)\}=\int\mu(d\alpha)\mathbb{E}\Phi'_\alpha(A_\alpha)$$

and hence with Taylor expansion

$$
\begin{aligned}
\sigma^2\mathbb{E}f'(W) = & \int \mu(d\alpha)\mathbb{E}\Phi'_\alpha(A_\alpha)\mathbb{E}f'(\Phi'(B^c_\alpha)) \\
& + \int \mu(d\alpha)\mathbb{E}\Phi'_\alpha(A_\alpha)\mathbb{E}\biggl\{\Phi'(B_\alpha)\int_0^1 f''(\Phi'(B^c_\alpha)+t\Phi'(B_\alpha))\,dt\biggr\} \\
=: & \, R_1 + R_2.
\end{aligned}
$$

Now, again by Taylor,

$$
\begin{aligned}
\mathbb{E}\{Wf(W)\} &= \int \mu(d\alpha)\bigl[\mathbb{E}f(\Phi'_\alpha(J)) - \mathbb{E}f(\Phi'(J))\bigr] \\
&= \int \mu(d\alpha)\bigl[\mathbb{E}f(\Phi'_\alpha(A^c_\alpha)) - \mathbb{E}f(\Phi'(A^c_\alpha))\bigr] \\
&\quad + \int \mu(d\alpha)\Bigl[\mathbb{E}\bigl\{\Phi'_\alpha(A_\alpha)f'(\Phi'_\alpha(A^c_\alpha))\bigr\} - \mathbb{E}\bigl\{\Phi'(A_\alpha)f'(\Phi'(A^c_\alpha))\bigr\}\Bigr] \\
&\quad + \int \mu(d\alpha)\biggl[\mathbb{E}\biggl\{\Phi'_\alpha(A_\alpha)^2\int_0^1(1-t)f''\bigl(\Phi'_\alpha(A^c_\alpha)+t\Phi'_\alpha(A_\alpha)\bigr)\,dt\biggr\} \\
&\qquad\qquad - \mathbb{E}\biggl\{\Phi'(A_\alpha)^2\int_0^1(1-t)f''\bigl(\Phi'(A^c_\alpha)+t\Phi'(A_\alpha)\bigr)\,dt\biggr\}\biggr] \\
&=: R_3 + R_4 + R_5
\end{aligned}
$$

and

$$
\begin{aligned}
R_4 = & \int \mu(d\alpha)\bigl[\mathbb{E}\bigl\{\Phi'_\alpha(A_\alpha)f'(\Phi'_\alpha(B^c_\alpha))\bigr\} - \mathbb{E}\bigl\{\Phi'(A_\alpha)f'(\Phi'(B^c_\alpha))\bigr\}\bigr] \\
& + \int \mu(d\alpha)\biggl[\mathbb{E}\biggl\{\Phi'_\alpha(A_\alpha)\Phi'_\alpha(B_\alpha\setminus A_\alpha)\int_0^1 f''\bigl(\Phi'_\alpha(B^c_\alpha)+t\Phi'_\alpha(B_\alpha\setminus A_\alpha)\bigr)\,dt\biggr\} \\
&\qquad - \mathbb{E}\biggl\{\Phi'(A_\alpha)\Phi'(B_\alpha\setminus A_\alpha)\int_0^1 f''\bigl(\Phi'(B^c_\alpha)+t\Phi'(B_\alpha\setminus A_\alpha)\bigr)\,dt\biggr\}\biggr] \\
=: & \, R_6 + R_7.
\end{aligned}
$$

Using (3.8)–(3.10), we see that $R_3 = 0$ and $R_1 = R_6$, hence

$$
\bigl|\mathbb{E}\bigl\{Wf(W) - \sigma^2 f'(W)\bigr\}\bigr| \leqslant |R_2| + |R_5| + |R_7|.
$$

With (2.15) we finally obtain

$$
\begin{aligned}
|R_2| &\leqslant \|\Delta\hat{g}\|\int \mu(d\alpha)\bigl|\mathbb{E}\Phi'_\alpha(A_\alpha)\bigr|\mathbb{E}\bigl\{|\Phi'(B_\alpha)|D^1\bigl[\mathscr{L}\bigl(\Phi(B^c_\alpha)\bigm|\Phi(B_\alpha)\bigr)\bigr]\bigr\}, \\
|R_5| &\leqslant \tfrac{1}{2}\|\Delta\hat{g}\|\int \mu(d\alpha)\Bigl[\mathbb{E}\bigl\{\Phi'_\alpha(A_\alpha)^2 D^1\bigl[\mathscr{L}\bigl(\Phi_\alpha(A^c_\alpha)\bigm|\Phi_\alpha(A_\alpha)\bigr)\bigr]\bigr\} \\
&\qquad + \mathbb{E}\bigl\{\Phi'(A_\alpha)^2 D^1\bigl[\mathscr{L}\bigl(\Phi(A^c_\alpha)\bigm|\Phi(A_\alpha)\bigr)\bigr]\bigr\}\Bigr], \\
|R_7| &\leqslant \|\Delta\hat{g}\|\int \mu(d\alpha)\Bigl[\mathbb{E}\bigl\{|\Phi'_\alpha(A_\alpha)\Phi'_\alpha(B_\alpha\setminus A_\alpha)|D^1\bigl[\mathscr{L}\bigl(\Phi_\alpha(B^c_\alpha)\bigm|\Phi_\alpha(A_\alpha),\Phi_\alpha(B_\alpha)\bigr)\bigr]\bigr\} \\
&\qquad + \mathbb{E}\bigl\{|\Phi'(A_\alpha)\Phi'(B_\alpha\setminus A_\alpha)|D^1\bigl[\mathscr{L}\bigl(\Phi(B^c_\alpha)\bigm|\Phi(A_\alpha),\Phi(B_\alpha)\bigr)\bigr]\bigr\}\Bigr].
\end{aligned}
$$

To obtain ϑ_2, just replace $\|\Delta g\|$ by $\|g\|$ and D^1 by D^2 in the above bounds . $\square$

Corollary 3.4. *Let Φ be a simple point process satisfying* (3.8)–(3.10). *If there is further a function $c_l(\alpha)$, such that for μ-almost every $\alpha \in J$ almost surely*

$$D^l\big[\mathscr{L}\big(\Phi(J) \mid \Phi|_{B_\alpha}\big)\big], D^l\big[\mathscr{L}\big(\Phi_\alpha(J) \mid \Phi_\alpha|_{B_\alpha}\big)\big] \leqslant c_l(\alpha), \qquad l = 1, 2, \tag{3.13}$$

then (3.12) *satisfies*

$$\begin{aligned}\vartheta_l(\alpha) \leqslant c_l(\alpha)\Big[&\big|\mathbb{E}\Phi'_\alpha(A_\alpha)\big|\mathbb{E}|\Phi'(B_\alpha)| + \tfrac{1}{2}\mathbb{E}\Phi'_\alpha(A_\alpha)^2 + \tfrac{1}{2}\mathbb{E}\Phi'(A_\alpha)^2 \\ &+ \mathbb{E}|\Phi'_\alpha(A_\alpha)\Phi'_\alpha(B_\alpha \setminus A_\alpha)| + \mathbb{E}|\Phi'(A_\alpha)\Phi'(B_\alpha \setminus A_\alpha)|\Big] \\ \leqslant c_l(\alpha)\Big[&1.5\mathbb{E}\big\{\Phi_\alpha(A_\alpha)\Phi_\alpha(B_\alpha)\big\} + 1.5\mathbb{E}\big\{\Phi(A_\alpha)\Phi(B_\alpha)\big\} \\ &+ 6\mu(A_\alpha)\mu(B_\alpha) + 4\mu(B_\alpha)\mathbb{E}\Phi_\alpha(B_\alpha)\Big].\end{aligned} \tag{3.14}$$

4. Applications

In what follows, we calculate only rough bounds, leaving much scope for improvement. In particular, we replace the moments in the estimates by almost sure bounds.

4.1. Exceedances of the r-scans process. We follow the notation of Dembo and Karlin (1992). Let $X_1, X_2, \dots, X_{n+r-1}$ be independent and identically distributed random variables with distribution function F. Define the r-scan process $R_i = \sum_{k=0}^{r-1} X_{i+k}$, $i = 1, 2, \dots, n$ and further $W_i^- = I[R_i \leqslant a]$ for $a \in \mathbb{R}$. We are interested in the number $N^- = \sum_{i=1}^n W_i^-$, that is the number of R_i not exceeding a. With $p = \mathbb{E}W_i^- = \mathbb{P}[R_1 \leqslant a]$, we have $\mathbb{E}N^- = np$ and

$$\sigma^2 = \operatorname{Var} W = np\Big(1 - p + 2\sum_{d=1}^{r-1}(1 - d/n)\psi(d)\Big), \tag{4.1}$$

where $\psi(d) = \mathbb{P}[R_{d+1} \leqslant a | R_1 < a] - p \geqslant 0$.

Poisson approximations for the r-scan process have been extensively studied by Dembo and Karlin (1992). Normal approximation has been considered by Dembo and Rinott (1996); in particular they show, that, for fixed r and a, N^- converges in the Kolmogorov metric to the normal distribution with rate $\mathrm{O}(n^{-1/2})$. In the next theorem we achieve the same rate in total variation, and also a rate for the corresponding local limit approximation.

Theorem 4.1. *Assume that F is continuous, $F(0) = 0$, and $0 \leqslant F(x) < F(y)$ for all $x < y$, and let $a > 0$ be fixed. Then, for all n such that $\sigma^2 > 1$,*

$$d_l\big(\mathscr{L}(N^- - np), \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2\rceil, 1/2 - t\big)\big) \leqslant C_l n^{-l/2}, \qquad l = 1, 2,$$

where the constants C_1 and C_2 are independent of n and can be extracted from the proof.

Proof. We apply Theorem 3.1 for $W = \sum_{i=1}^{n} \xi_i = \sum_{i=1}^{n}(W_i^- - p)$. We can set

$$A_i = \{i-r+1,\dots,i+r-1\} \cap \{1,\dots,n\},$$
$$B_i = \{i-2r+2,\dots,i+2r-2\} \cap \{1,\dots,n\}.$$

Then, as $|A_i| \leqslant 2r-1$, $|B_i| \leqslant 4r-3$ and $|B_i \setminus A_i| \leqslant 2r-2$, the following rough bounds are obvious:

$$\mathbb{E}|\xi_i\eta_i^2| \leqslant (2r-1)^2, \quad \mathbb{E}|\xi_i\eta_i(\tau_i-\eta_i)| \leqslant (2r-1)(2r-2),$$
$$|\mathbb{E}\xi_i\eta_i|\mathbb{E}|\tau_i| \leqslant (2r-1)(4r-3),$$

thus

$$\vartheta_{l,i} \leqslant c_{l,i}\big(16r^2-20r+6\big) \tag{4.2}$$

Consider now the block $B_1 = \sum_{i=1}^{3r-2} W_i^-$, and assume that the values $\partial B_1 = (X_1,\dots,X_{r-1})$ and $\partial B_2 = (X_{3r-1},\dots,X_{4r-2})$ are given. Define the events

$$\mathcal{A} := \big\{a/r < X_r,\dots,X_{2r-2},X_{2r+1},\dots,X_{3r-2} \leqslant a(r+1)/r^2,\ 0 < X_{2r} \leqslant a/(2r^2)\big\}$$
$$\mathcal{A}_0 := \big\{a/r < X_{2r-1} \leqslant a(r+1)/r^2\big\}, \quad \mathcal{A}_1 := \big\{0 < X_{2r-1} \leqslant a/(2r^2)\big\}.$$

Due to the conditions on F and independence it is clear that $p_j := \mathbb{P}[\mathcal{A}\cap\mathcal{A}_j] > 0$ for $j=0,1$. Note now that

$$R_r = \sum_{i=r}^{2r-1} X_i > a \quad \text{on } \mathcal{A}\cap\mathcal{A}_0, \qquad R_r < a \quad \text{on } \mathcal{A}\cap\mathcal{A}_1.$$

Note further that $R_s < a$ for all $s = r+1,\dots,2r-1$ on $\mathcal{A}\cap(\mathcal{A}_0\cup\mathcal{A}_1)$. Hence

$$\sum_{i=r}^{2r-1} W_i^- = r-1 \quad \text{on } \mathcal{A}\cap\mathcal{A}_0, \qquad \sum_{i=r}^{2r-1} W_i^- = r \quad \text{on } \mathcal{A}\cap\mathcal{A}_1.$$

It easy to see now by a coupling argument that

$$\tfrac{1}{2}D^1\big(\mathscr{L}(B_1)\big) \leqslant 1-(p_0\wedge p_1) < 1.$$

Noting that by sequentially stringing together blocks like B_1, we can have $m := \lfloor n/(3r-2)\rfloor$ such blocks, which are independent given all the borders ∂B_i. Furthermore, for every i, the R_j in B_i depend on the X_k of at most two such blocks. Therefore, defining $Z = (\partial B_1,\dots,\partial B_m)$ and using (5.11) and (5.12),

$$D^1\big(\mathscr{L}(W|\partial B_i, i=1,\dots,m)\big) \leqslant \frac{2}{\big(\min\{1/2,p_0,p_1\}(m-2)\big)^{1/2}} =: c_{1,i},$$
$$D^2\big(\mathscr{L}(W|\partial B_i, i=1,\dots,m)\big) \leqslant \frac{8}{\min\{1/2,p_0,p_1\}(m-4)_+} =: c_{2,i}.$$

Clearly, $c_{l,i} = \mathrm{O}(n^{-l/2})$. Hence, putting this, (4.1) and (4.2) into (3.1), the theorem follows. $\square$

4.2. Matérn hard-core process type I. We approximate the total number of points of the Matérn hard-core process type I introduced by Matérn (1960). We use rectangular instead of the usual circular neighborhoods. Let Φ be the process on the d-dimensional cube $J = [0,1)^d \subset \mathbb{R}^d$ defined as

$$\Phi(B) = \sum_{i=1}^{\tau} I[X_i \in B] I[X_j \notin K_r(X_i) \text{ for all } j = 1, \dots, \tau,\ j \neq i],$$

where $\tau \sim \mathrm{Po}(\lambda)$ and $\{X_i;\ i \in \mathbb{N}\}$ is a sequence of independent and uniformly distributed random variables on J and where, for $x = (x_1, \dots, x_d) \in J$ and $r > 0$, $K_r(x)$ denotes the d-dimensional closed cube with center x and side length r. To avoid edge effects, we treat J as a d-dimensional torus, thus identifying any point outside J by the point in J which results in coordinate-wise shifting by 1. The process Φ is thus a thinned Poisson point process with rate λ having all points deleted which contain another point in their K_r neighborhood. For the mean measure μ of Φ we obtain

$$\frac{d\mu(x)}{dx} = \lambda \mathrm{e}^{-c}. \tag{4.3}$$

We are now interested in the distribution of $\Phi(B)$ when r is small an λ large.

Theorem 4.2. *Put $W := \Phi(J) - \mu(J)$ and let $a > 0$ be a fixed real number. Then, for every λ and r such that $\lambda r^d = a$ and $\sigma^2 := \operatorname{Var} W > 1$,*

$$d_l\big(\mathscr{L}(\Phi(J) - \mu(J)), \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2 \rceil, 1/2 - t\big)\big) \leqslant C_l \lambda^{-l/2}, \qquad l = 1, 2,$$

for constants C_1 and C_2 which are independent of λ and can be extracted from the proof.

Proof. We apply Corollary 3.4. We can take $A_x = K_{2r}(x)$ and $B_x = K_{4r}(x)$ and check that the conditions (3.8)–(3.10) are fulfilled. Some calculations show that the reduced second factorial moment measure M satisfies

$$\frac{dM(x)}{dx} = \begin{cases} 0 & \text{if } x \in K_r(0), \\ \lambda^2 \mathrm{e}^{-\lambda |K_r(0) \cup K_r(x)|} & \text{if } x \in K_{2r}(0) \setminus K_r(0), \\ \lambda^2 \mathrm{e}^{-2a} & \text{if } x \notin K_{2r}(0), \end{cases}$$

compare with Daley and Vere-Jones (1988, pp. 367, 373). Thus, $M(J) \geqslant \lambda^2 \mathrm{e}^{-2a}(1 - r^d)$ and

$$\sigma^2 = \lambda \mathrm{e}^{-a} + M(J) - \mu(J)^2 \geqslant \lambda \mathrm{e}^{-a}(1 - a\mathrm{e}^{-a}). \tag{4.4}$$

Since we can have at most 7^d points of Φ in B_x, we obtain from (3.14) the rough estimate

$$\vartheta_l(x) \leqslant 26 \cdot 7^d c_l(x), \tag{4.5}$$

where $c_l(\cdot)$ is as in (3.13). To estimate $c_l(x)$ write $K_r = K_r(0)$. We have

$$\mathbb{P}\big[\Phi(K_r) = 0 \;\big|\; \Phi_{K_{7r} \setminus K_{5r}}\big] \geqslant \mathrm{Po}\big(\lambda |K_r|\big)\{0\} = \mathrm{e}^{-a} =: p_0,$$

$$\mathbb{P}\big[\Phi(K_r) = 1 \;\big|\; \Phi_{K_{7r} \setminus K_{5r}}\big] \geqslant \mathrm{Po}\big(\lambda |K_{3r} \setminus K_r|\big)\{0\} \cdot \mathrm{Po}\big(\lambda |K_r|\big)\{1\} = a\mathrm{e}^{-3^d a} =: p_1.$$

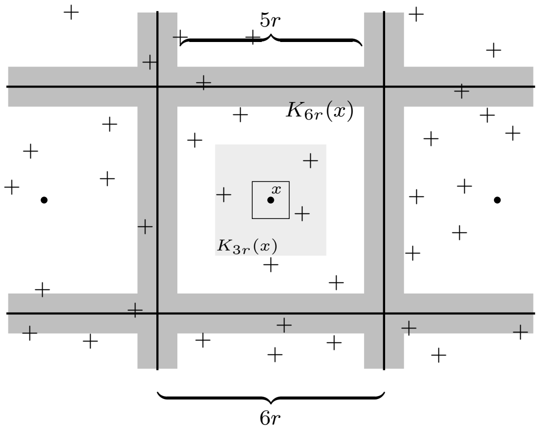

FIG 1. *Matérn hard-core process type I: Given that the process Φ is known on the borders $\cup_{l\in\mathcal{M}} K_{6r}(x_l)\setminus K_{5r}(x_l)$ (grey area), the boxes $\Phi|_{K_{6r}(x_l)}$, $l\in\mathcal{M}$, are independent.*

Hence, by a coupling argument,

$$\tfrac{1}{2}D^1\big(\mathscr{L}(\Phi(K_{6r})|\Phi_{K_{7r}\setminus K_{5r}})\big) \leqslant 1-(p_0\wedge p_1)<1. \tag{4.6}$$

Let now x be arbitrary. Divide the space J into boxes of side length $6r$, centered around x (see Figure 1). With $m := \lfloor 1/(6r)\rfloor$, we can have m^d such boxes plus a remainder. Denote this remainder by J^R and denote by x_l, $l\in\{1,\dots,m\}^d =: \mathcal{M}$ the centers of the boxes where $x_{1,\dots,1}=x$. Note now that, given Φ on all the borders $K_{6r}(x_l)\setminus K_{5r}(x_l)$, $l\in\mathcal{M}$ (grey area in Figure 1), the random variables $\Phi(K_{6r}(x_l))$, $l\in\mathcal{M}$, are independent and satisfy inequality (4.6). Furthermore, $\Phi|_{J\setminus K_{6r}(x)}$ is independent of $\Phi|_{B_x}$, and therefore, defining $Z=\big((\Phi|_{K_{6r}(x_l)\setminus K_{5r}(x_l)})_{l\in\mathcal{M}},\Phi|_{J^R}\big)$ and using (5.11) and (5.12), we obtain

$$D^1\big(\mathscr{L}\big(\Phi(J)\bigm|\Phi|_{B_x}\big)\big) \leqslant \frac{2}{\big(\min\{1/2,p_0,p_1\}(m^d-2)_+\big)^{1/2}} =: c_1(x), \tag{4.7}$$

$$D^2\big(\mathscr{L}\big(\Phi(J)\bigm|\Phi|_{B_x}\big)\big) \leqslant \frac{8}{\min\{1/2,p_0,p_1\}(m^d-3)_+} =: c_2(x). \tag{4.8}$$

Noting that almost surely $\mathscr{L}\big[\Phi_x\big(J\setminus K_{6r}(x)\big)\bigm|\Phi_x|_{B_x}\big]=\mathscr{L}\big[\Phi\big(J\setminus K_{6r}(x)\big)\big]$, we see that (4.7) and (4.8) hold also for Φ_x, thus $c_l(x)$ satisfies (3.13). Now, recalling that $a=\lambda r^d$ is constant, we have $c_l(x)=\mathrm{O}(\lambda^{-l/2})$. Hence, putting this and (4.3)–(4.5) into (3.11), the theorem follows. $\square$

5. Appendix

5.1. Properties of the solution to the Stein equation.

Lemma 5.1. *For any indicator function $h(z) = I[z \in A]$, $z \in \mathbb{Z}$, $A \subset \mathbb{Z}$, the solution $g = g_h$ to the Stein equation* (2.3) *satisfies*

$$\|g\| \leqslant 1 \wedge (npq)^{-1/2}. \tag{5.1}$$

Proof. We apply the generator method introduced by Barbour (1988). For any function $f : \{0, \dots, n\} \to \mathbb{R}$, define

$$(\mathcal{A}f)(z) = \big(\mathcal{B}(-\Delta f)\big)(z) = qzf(z-1) - \big(qz + p(n-z)\big)f(z) + p(n-z)f(z+1), \tag{5.2}$$

which is the infinitesimal generator of a pure jump Markov process. A solution g to (2.3) is now given through

$$\psi(z) = -\int_0^\infty \mathbb{E}\big\{h\big(Y_z(t)\big) - h(Y)\big\}\,dt, \qquad \text{for } z \in \{0, \dots, n\},$$

and $g(z) = -\Delta\psi(z)$ for $z \in \{0, \dots, n-1\}$ and $g(z) = 0$ else, where Y_z is a Markov process with generator $\mathcal{A}$ starting at point z, and Y is a random variable having the stationary distribution $\mathrm{Bi}(n, p)$. Now, we have for $z \in \{0, \dots, n-1\}$,

$$\Delta\psi(z) = \int_0^\infty \mathbb{E}\big\{h\big(Y_z(t)\big) - h\big(Y_{z+1}(t)\big)\big\}\,dt. \tag{5.3}$$

We now fix z and construct a coupling of Y_z and Y_{z+1} to bound (5.3). Let thereto $X_k^{(i)}(t)$, $k \in \{1, \dots, n\}$, $i \in \{0, 1\}$, be independent Markov processes with state space $\{0, 1\}$, starting in point i and having jump rate p if the process is in 0 and q otherwise. It is easy to see by the Kolmogorov differential equations that

$$X_k^{(1)}(t) \sim \mathrm{Be}(p + q\mathrm{e}^{-t}), \qquad X_k^{(0)}(t) \sim \mathrm{Be}(p - p\mathrm{e}^{-t}) \tag{5.4}$$

where $\mathrm{Be}(p)$ denotes the Bernoulli distribution with success probability p. Let τ be the minimum of the first jump times of the two processes $X_{z+1}^{(0)}$ and $X_{z+1}^{(1)}$, and define a new process

$$X(t) = \begin{cases} X_{z+1}^{(1)} & \text{if } \tau > t, \\ X_{z+1}^{(0)} & \text{if } \tau \leqslant t, \end{cases}$$

describing the well-known Doeblin coupling. Then, let

$$Y_z = \sum_{k=1}^{z} X_k^{(1)} + \sum_{k=z+1}^{n} X_k^{(0)}, \qquad Y_{z+1} = Y_z - X_{z+1}^{(0)} + X(t), \tag{5.5}$$

and one proves that Y_z and Y_{z+1} are Markov processes with generator (5.2). Hence, we can write (5.3) as

$$-\Delta\psi(z) = \int_0^\infty \mathrm{e}^{-t}\mathbb{E}\big\{\Delta h(Y_z)\big\}\,dt, \tag{5.6}$$

since τ is exponentially distributed with rate 1. The bound $\|g\| \leqslant 1$ is now immediate from (5.6), thus we may assume that $npq > 1$. Note that, from (5.4) and (5.5),

$$\mathscr{L}(Y_z) = \mathrm{Bi}(z, p + q\mathrm{e}^{-t}) * \mathrm{Bi}(n - z, p - p\mathrm{e}^{-t}),$$

and hence, from Barbour and Jensen (1989, Lemma 1),

$$\begin{aligned} D^1\bigl(\mathscr{L}(Y_z)\bigr) &\leqslant \mathrm{Var}(Y_k)^{-1/2} \\ &\leqslant \bigl(z(p + q\mathrm{e}^{-t})(q - q\mathrm{e}^{-t}) + (n - z)(p - p\mathrm{e}^{-t})(q + p\mathrm{e}^{-t})\bigr)^{-1/2} \\ &\leqslant \bigl(npq(1 - \mathrm{e}^{-t})\bigr)^{-1/2}. \end{aligned} \tag{5.7}$$

Note also that for $\tilde{h} := h - 1/2$

$$\bigl|\mathbb{E}\bigl\{\Delta h(Y_z)\bigr\}\bigr| = \bigl|\mathbb{E}\bigl\{\Delta \tilde{h}(Y_z)\bigr\}\bigr| \leqslant D^1\bigl(\mathscr{L}(Y_z)\bigr)/2. \tag{5.8}$$

Thus, applying (5.8) on (5.6) and using (5.7),

$$\bigl|\Delta\psi\bigr| \leqslant \int_0^s \mathrm{e}^{-t}\,dt + \frac{1}{2\sqrt{npq}} \int_s^\infty \frac{\mathrm{e}^{-t}}{\sqrt{1 - \mathrm{e}^{-t}}}\,dt.$$

Choosing $s = -\ln\bigl(1 - (npq)^{-1}\bigr)$ and computing the integrals proves the lemma. □

5.2. Change of the success probabilities.

Lemma 5.2. *For every $n \in \mathbb{N}$, $0 < p < 1$ and $-(1 - p) < t < p$*

$$d_{\mathrm{TV}}\bigl(\mathrm{Bi}(n, p - t), \mathrm{Bi}(n, p)\bigr) \leqslant |t| \biggl(\frac{\sqrt{n}}{\sqrt{pq}} + \frac{p - t}{pq} + \frac{\sqrt{(p - t)(q + t)}}{pq\sqrt{n}}\biggr)$$

$$d_{\mathrm{loc}}\bigl(\mathrm{Bi}(n, p - t), \mathrm{Bi}(n, p)\bigr) \leqslant |t| \biggl(\frac{1 + p - t}{pq} + \frac{\sqrt{(p - t)(q + t)}}{pq\sqrt{n}}\biggr)$$

Proof. We use Stein's method. If $W \sim \mathrm{Bi}(n, p - t)$, we obtain from (2.1) and (2.2)

$$\mathbb{E}\bigl\{(1 - p)Wg(W - 1) - p(n - W)g(W)\bigr\} = \mathbb{E}\bigl\{tW\Delta g(W - 1) - tng(W)\bigr\}$$

for every bounded function $g \in F(\mathbb{Z})$. The left side is just the Stein operator for $\mathrm{Bi}(n, p)$ hence, taking $g = g_A$ obtained by solving (2.3) for $\mathrm{Bi}(n, p)$, with the bounds (2.4) and (5.1) the d_{TV}-bound follows, noting also that $\mathbb{E}|W| \leqslant |\mathbb{E}W| + \sqrt{\mathrm{Var}\,W}$. With the remark after (2.4), the d_{loc}-bound is proved. □

5.3. Smoothing properties of independent random variables. In several parts of this paper, we have the situation that we need to estimate $D^m(U)$, $m = 1, 2$, for some integer valued random variable U, being a sum of some other random variables. If the U is a sum of independent random variables, we can proceed as follows. Assume that $U = \sum_{i=1}^n X_i$, where the X_i are independent. Defining $v_i = \min\{\frac{1}{2}, 1 - \frac{1}{2}D^1(X_i)\}$ and $V = \sum_i v_i$ we obtain from Barbour and Xia (1999, Proposition 4.6) the bound

$$D^1(U) \leqslant \frac{2}{V^{1/2}}. \tag{5.9}$$

Define further $v^* = \max_i v_i$. Now it is always possible to write $U = U^{(1)} + U^{(2)}$ in such a way that the analogously defined numbers $V^{(1)}$ and $V^{(2)}$ satisfy $V^{(k)} \geqslant V/2 - v^*$, $k = 1, 2$. Using (1.2) and (5.9), we obtain

$$D^2(U) \leqslant D^1\big(U^{(1)}\big) D^1\big(U^{(2)}\big) \leqslant \frac{4}{\big(V^{(1)}V^{(2)}\big)^{1/2}} \leqslant \frac{8}{(V-2v^*)_+}. \tag{5.10}$$

5.4. Smoothing properties of conditional independent random variables. In most applications, U is a sum of dependent summands and we can not apply (5.9) and (5.10) directly. However, assuming that there is a random variable Z on the same probability space as U such that $\mathscr{L}(U|Z=z)$ can be represented as a sum of independend summands, say $X_i^{(z)}$, $i = 1, \dots, n_z$, for each z that Z can attain, we can still apply (5.9) and (5.10), and we obtain

$$D^1(U) \leqslant \mathbb{E}\big\{\mathbb{E}[D^1(U)|Z]\big\} \leqslant \mathbb{E}\Big\{\frac{2}{V_Z^{1/2}}\Big\}, \tag{5.11}$$

$$D^2(U) \leqslant \mathbb{E}\big\{\mathbb{E}[D^1(U)|Z]\big\} \leqslant \mathbb{E}\Big\{\frac{8}{(V_Z - 2v_Z^*)_+}\Big\}, \tag{5.12}$$

where, for each z, V_z and v_z^* are the corresponding values as defined in subsection 5.3 with respect to the $X_i^{(z)}$.

6. Appendix

We now give a generalization of Theorem 3.1. The proof is omitted, because it runs analogously to the proof of Theorem 3.1; see also Barbour et al. (1989).

Suppose that a random variable W satisfies Assumptions G and assume that there are sets $K_i \subset J$, $i \in I$, and square integrable random variables Z_i, Z_{ik} and V_{ik}, $k \in K_i$ and $i \in I$, as follows:

$$W = W_i + Z_i,\ i \in I, \text{ where } W_i \text{ is independent of } \xi_i, \tag{6.1}$$

$$Z_i = \sum_{k \in K_i} Z_{ik}, \tag{6.2}$$

$$W_i = W_{ik} + V_{ik},\ i \in I,\ k \in K_i, \tag{6.3}$$

where W_{ik} is independent of the pair (X_i, Z_{ik}).

Theorem 6.1. *With W as above,*

$$d_l\big(\mathscr{L}(W), \widehat{\mathrm{Bi}}\big(\lceil 4\sigma^2 \rceil, 1/2 - t\big)\big) \leqslant \sigma^{-2}\Big(\sum_{i \in I} \vartheta_{l,i} + 1.75\Big), \qquad l = 1, 2, \tag{6.4}$$

where

$$\begin{aligned} \vartheta_{l,i} = {} & \frac{1}{2}\mathbb{E}\big\{|\xi_i| Z_i^2\, D^l\big(\mathscr{L}(W_i|\xi_i, Z_i)\big)\big\} \\ & + \sum_{k \in K_i} \mathbb{E}\big\{\big|\xi_i Z_{ik} V_{ik}\big| D^l\big(\mathscr{L}(W_{ik}|\xi_i, Z_{ik}, V_{ik})\big)\big\} \\ & + \sum_{k \in K_i} \big|\mathbb{E}\{\xi_i Z_{ik}\}\big|\, \mathbb{E}\big\{|Z_i + V_{ik}| D^l\big(\mathscr{L}(W_{ik}|Z_i, V_{ik})\big)\big\}. \end{aligned} \tag{6.5}$$

Acknowledgments

I thank A. D. Barbour, D. Schuhmacher and B. Nietlispach for many helpful discussions.

References

A. D. Barbour (1988). Stein's Method and Poisson Process Convergence. *J. Appl. Probab.* **25A**, 175–184.

A. D. Barbour and J. L. Jensen (1989). Local and Tail Approximation Near the Poisson Limit. *Scand. J. Statist.* **16**, 75–87.

A. D. Barbour and A. Xia (2001). The number of two-dimensional maxima. *Adv. in Appl. Probab.* **33**, 727–750.

A.D. Barbour and V. Čekanavičius (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30**, 509–545.

A.D. Barbour and A. Xia (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3**, 131–159.

A.D. Barbour, M. Karoński, and A. Ruciński (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47**, 125–145.

A.D. Barbour, L. Holst, and S. Janson (1992). Poisson Approximation. Oxford University Press.

T. C. Brown and A. Xia (2001). Stein's method and birth-death processes. *Ann. Probab.* **29**, 1373–1403.

V. Čekanavičius (1997). Asymptotic expansions in the exponent: a compound Poisson approach. *Adv. in Appl. Probab.* **29**, 374–387.

V. Čekanavičius (1998). Poisson approximations for sequences of random variables. *Statist. Probab. Lett.* **39**, 101–107.

V. Čekanavičius and P. Vaĭtkus (2001). Centered Poisson approximation by the Stein method. *Lithuanian Math. J.* **41**, 319–329.

L. H. Y. Chen and Q.-M. Shao (2004). Normal approximation under local dependence. *Ann. Probab.* **32**, 1985–2028.

D. J. Daley and D. Vere-Jones (1988). An introduction to the theory of point processes. Springer Series in Statistics. Springer-Verlag, New York.

A. Dembo and S. Karlin (1992). Poisson approximations for r-scan processes. *Ann. Appl. Probab.* **2**, 329–357.

A. Dembo and Y. Rinott (1996). Some examples of normal approximations by Stein's method. In *Random Discrete Structures*, *IMA Vol. Math. Appl.* **76**, 25–44. Springer, New York.

W. Ehm (1991). Binomial approximation to the Possion Binomial distribution. *Statist. Probab. Lett.* **11**, 7–16, 1991.

L. Le Cam (1965). On the distribution of sums of independent random variables. In *Proc. Internat. Res. Sem., Statist. Lab., Univ. California, Berkeley, Calif.*, 179–202. Springer-Verlag, New York.

B. Matérn (1960). Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden Fran Statens Skogsforskningsinstitut*, Band 49, Nr. 5. Stockholm.

G. Reinert (1998). Coupling for normal approximations with Stein's method. *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **41**, 193–207.

A. Röllin (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.

B. Roos (2003). Kerstan's method for compound Poisson approximation. *Ann. Probab.* **31**, 1754–1771.

C. Stein (1986). Approximate computation of expectation. *IMS Lecture Notes Monogr. Ser.* **7**.

submitted to the Annals of Applied Probability

TRANSLATED POISSON APPROXIMATION USING EXCHANGEABLE PAIR COUPLINGS

By Adrian Röllin*

It is shown that the method of exchangeable pairs introduced by Stein (1986) for normal approximation can effectively be used for translated Poisson approximation. Introducing an additional smoothness condition, one can obtain approximation results in total variation and also in a local limit metric. The result is applied in particular to the anti-voter model on finite graphs as analysed by Rinott and Rotar (1997), obtaining the same rate of convergence, but now for a stronger metric.

1. Introduction

Let W be a random variable with $\mathbb{E}W = \mu$ and $\operatorname{Var} W = \sigma^2$. Stein (1986) introduced a method (which is commonly called the *exchangeable pair approach*) to approximate $W_c := (W-\mu)/\sigma$ by the standard normal distribution; Rinott and Rotar (1997) then generalised the result and successfully applied it to weighted U-statistics and the antivoter model. Their results imply convergence in the Kolmogorov and even in some stronger metrics; however, they do not to provide approximations in the total variation metric or prove local limit like results.

We will consider such results in this paper in the special case, in which the sum W is integer valued, the most common situation being the one where W is a sum of random indicators. As the total variation between W and the normal distribution will always be 1, we will instead use a translated Poisson distribution, matching the first two moments of W as well as possible. Note that we will consider the approximation of the unstandardised variable W and assume that $\sigma^2 \to \infty$, so there is actually no convergence taking place. We will also consider a metric from which local limit approximations can be obtained.

Recall the setting of Stein (1986) and Rinott and Rotar (1997). A pair of random variables (W, W') is called exchangeable, if $\mathscr{L}(W, W') = \mathscr{L}(W', W)$. Assume now that there is a positive number $\lambda < 1$ and a random variable R such that

$$\mathbb{E}^W(W' - \mu) = (1-\lambda)(W-\mu) + R, \tag{1.1}$$

holds, where $\mathbb{E}^W$ denotes the conditional expectation with respect to W. Of course, one can always find R to satisfy (1.1), so R must be thought of as being small for the approximation to be successful. Note that (1.1) implies $\mathbb{E}R = 0$.

If the pair (W, W') can be chosen such that condition (1.1) holds and $\mathbb{E}^W(W'-W)^2$ does not fluctuate too much, convergence of W_c to the standard normal distribution

*Partially supported by Schweizerischer Nationalfondsprojekt 20-107935/1

will follow in the Kolmogorov metric. As the behaviour of the difference $W' - W$ is mainly responsible for the quality of the approximation, it is an obvious starting point to introduce a smoothness condition, to make sure that the local perturbations of W are not too strong. Recall that Rinott and Rotar (1997) propose to choose W and W' as two successive steps of a reversible Markov chain with stationary distribution $\mathscr{L}(W)$. Then, condition (1.1) states that a particle on $\mathbb{Z}$ obeying the transition rules of such a Markov chain is forced to have a linear drift to the centre. Now $\mathbb{E}^{W=k}(W' - W)^2$ is the average of the squared jump size of the Markov chain if the particle is in k, so that for a good normal approximation, the average jump size of the particle has to be about the same wherever it is. It is now clear that, under these conditions, the particle may still behave irregularly on a local scale, for instance, the particle could still make only jumps of size two and thus stay on the odd or even integers, such that an approximation with a distribution on $\mathbb{Z}$ with span 1 will not be successful in total variation.

Thus, in addition to (1.1), we assume further that

$$W' - W \in \{-1, 0, +1\}, \tag{1.2}$$

and we will see that this seems to be an appropriate condition. Note that under condition (1.2) the corresponding Markov chain does not need to be reversible for (W, W') to be a exchangeable pair; see Lemma 1.1 of Rinott and Rotar (1997).

Condition (1.2) is in sharp contrast to other approaches using Stein's method for the translated Poisson distribution such as Čekanavičius and Vaĭtkus (2001), Röllin (2005) or Barbour and Lindvall (to appear), where an embedded sum of independent random variables within W is used for an explicit smoothing argument; in contrast, the smoothing effect of (1.2) will enter only implicitly into the proof of the main result. Chatterjee et al. (2005) use the same condition to obtain Poisson approximation results with the exchangeable pair approach.

As we are restricted to the integers, we cannot arbitrarily shift a Poisson distribution with a given variance to fit the mean, so some care is needed here. We say that an integer valued random variable Y has a *translated Poisson distribution* with parameters μ and σ^2 and write

$$\mathscr{L}(Y) = \mathrm{TP}(\mu, \sigma^2)$$

if $\mathscr{L}(Y - \mu + \sigma^2 + \gamma) = \mathrm{Po}(\sigma^2 + \gamma)$ where $\gamma = \langle \mu - \sigma^2 \rangle$ and $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x; in particular $\mathrm{TP}(\sigma^2, \sigma^2) = \mathrm{Po}(\sigma^2)$. So, approximating W with $\mathrm{TP}(\mu, \sigma^2)$, we can fit the mean exactly, but note that for the variance we have $\sigma^2 \leqslant \operatorname{Var} Y = \sigma^2 + \gamma \leqslant \sigma^2 + 1$. This will, however, cause no further problems, as the order of error of this mismatch is $\mathrm{O}(\sigma^{-2})$.

Throughout the paper, we shall be concerned with two metrics for probability distributions, the total variation metric d_{TV} and the local limit metric d_{loc}, where, for two probability distributions P and Q given by the point probabilities $\{p_k, k \in \mathbb{Z}\}$

and $\{q_k, k \in \mathbb{Z}\}$ respectively,

$$\begin{aligned} d_{\mathrm{TV}}\bigl(P,Q\bigr) &:= \sup_{A\subset\mathbb{Z}}\bigl|P(A)-Q(A)\bigr| = \frac{1}{2}\sum_{k\in\mathbb{Z}}|p_k-q_k|,\\ d_{\mathrm{loc}}\bigl(P,Q\bigr) &:= \sup_{k\in\mathbb{Z}}|p_k-q_k|. \end{aligned}$$

2. Main results

Theorem 2.1. *Assume that* (W, W') *is an exchangeable pair with values on the integers and which satisfies* (1.1) *and* (1.2). *Then, with* $S = S(W) = \mathbb{P}[W' = W+1|W]$ *and* $q_{\max} = \max_{k\in\mathbb{Z}} \mathbb{P}[W = k]$,

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) \leqslant \frac{\sqrt{\operatorname{Var} S}}{\lambda\sigma^2} + \frac{2\sqrt{\operatorname{Var} R}}{\lambda\sigma} + \frac{2}{\sigma^2}, \tag{2.1}$$

$$\begin{aligned} d_{\mathrm{loc}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) \leqslant{}& \frac{2\sqrt{q_{\max}\operatorname{Var} S}}{\lambda\sigma^2} + \frac{2q_{\max}\sqrt{\operatorname{Var} R}}{\lambda\sigma}\\ &+ \frac{\sqrt{\operatorname{Var} R}}{\lambda\sigma^2} + \frac{2}{\sigma^2}. \end{aligned} \tag{2.2}$$

Before proving Theorem 2.1, we give a short introduction into Stein's method for distributional approximation. The starting point for translated Poisson approximation is the Stein-Chen method for the Poisson distribution as presented in detail by Barbour et al. (1992).

Let W be an integer valued random variable with expectation μ and variance $\sigma^2 > 0$, and let $s = \lfloor \mu - \sigma^2 \rfloor$ and $\gamma = \langle \mu - \sigma^2 \rangle$ where $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that, if $Y \sim \mathrm{TP}(\mu,\sigma^2)$, then $Y - s \sim \mathrm{Po}(\sigma^2+\gamma)$. Let $\mathcal{A}g(j) = (\sigma^2+\gamma)g(j+1) - jg(j)$ be the usual Stein operator for the Poisson distribution with mean $\sigma^2+\gamma$, and for $A \subset \mathbb{Z}_+ := \{0,1,2,\dots\}$ let $g_A : \mathbb{Z} \to \mathbb{R}$ be the solution of

$$i)\ \ g(j) = 0 \text{ for all } j \leqslant 0, \tag{2.3}$$

$$ii)\ \ \mathcal{A}g(j) = I[j \in A] - \mathrm{Po}(\sigma^2+\gamma)\{A\} \text{ for all } j \geqslant 0. \tag{2.4}$$

We can thus bound the total variation distance as

$$\begin{aligned} d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) &= d_{\mathrm{TV}}\bigl(\mathscr{L}(W-s), \mathrm{Po}(\sigma^2+\gamma)\bigr)\\ &= \sup_{B\subset\mathbb{Z}}\bigl|\mathbb{E}I[W-s\in B] - \mathrm{Po}(\sigma^2+\gamma)\{B\}\bigr|\\ &\leqslant \sup_{A\subset\mathbb{Z}_+}\bigl|\mathbb{E}\mathcal{A}g_A(W-s)\bigr| + \mathbb{P}[W-s<0], \end{aligned} \tag{2.5}$$

The last terms in (2.5) can be bounded using Chebyshev's inequality as

$$\mathbb{P}[W-s<0] = \mathbb{P}\bigl[W-\mu < -(\sigma^2+\gamma)\bigr] \leqslant \mathbb{P}\bigl[|W-\mu| > \sigma^2+\gamma\bigr] \leqslant \frac{1}{\sigma^2}. \tag{2.6}$$

From (Barbour et al., 1992, Lemma 1.1.1) we have the well-known bounds on the supremum norm of g_A,

$$\|g_A\| \leqslant (\sigma^2+\gamma)^{-1/2} \leqslant \sigma^{-1}, \quad \|\Delta g_A\| \leqslant (\sigma^2+\gamma)^{-1} \leqslant \sigma^{-2}, \tag{2.7}$$

where $\Delta g_A(j) := g_A(j+1) - g_A(j)$. If $A = \{k\}$ for some $k \in \mathbb{Z}$ we even have

$$\|g_{\{k\}}\| \leqslant \sigma^{-2}. \tag{2.8}$$

For the proof of the results in the d_{loc} metric, we will also need the following non-standard but simple result.

Lemma 2.2. *Let g_i be the solution of* (2.3)–(2.4) *for $A = \{i\}$. Then*

$$\sum_k |\Delta g_i(k)| \leqslant 2\sigma^{-2}, \qquad \sum_k \big(\Delta g_i(k)\big)^2 \leqslant 4\sigma^{-4}. \tag{2.9}$$

Proof. Recall from (Barbour et al., 1992, Proof of Lemma 1.1.1) that $g_i(k)$ is negative and decreasing in $0 \leqslant k \leqslant i$ and positive and decreasing in $k > i$ with the only positive jump in i satisfying

$$|\Delta g_i(i)| \leqslant (\sigma^2+\gamma)^{-1} \leqslant \sigma^{-2}.$$

From this, it is easy to see that the first bound of (2.9) holds and the second bound is then immediate. □

With $\tilde{g}_A(j) := g_A(j-s)$ we can rewrite the Stein operator $\mathcal{A}$ as

$$\begin{aligned}\mathcal{A}g_A(W-s) &= (\sigma^2+\gamma)g_A(W-s+1) - (W-s)g_A(W-s)\\ &= \sigma^2\Delta\tilde{g}_A(W) - (W-\mu)\tilde{g}_A(W) + \gamma\Delta\tilde{g}_A(W).\end{aligned} \tag{2.10}$$

The bounds on $\tilde{g}_A$ are of course the same as on g_A in (2.7)–(2.9). Thus, the last term in (2.10) is easily bounded by

$$\big|\mathbb{E}\{\gamma\Delta\tilde{g}_A(W)\}\big| \leqslant \gamma\sigma^{-2} \leqslant \sigma^{-2}. \tag{2.11}$$

Inserting (2.10) into (2.5) and invoking the bounds (2.6) and (2.11) we obtain

$$\begin{aligned}&d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\big)\\ &\qquad\leqslant \sup_{A\subset\mathbb{Z}_+}\big|\mathbb{E}\big\{\sigma^2\Delta\tilde{g}_A(W) - (W-\mu)\tilde{g}_A(W)\big\}\big| + 2\sigma^{-2};\end{aligned} \tag{2.12}$$

the same estimate holds for d_{loc} but with the supremum taken only over the sets $A = \{i\}$ for $i \in \mathbb{Z}_+$.

Proof of Theorem 2.1. We only have to bound the supremum in (2.12). Stein (1986) showed, that, if F satisfies $F(w,w') = -F(w',w)$ for all w and w', exchangeability implies $\mathbb{E}F(W,W') = 0$. Define the random variable $D := W'-W$ and the function

$F(w,w') := (w'-w)\big(g(w')+g(w)\big)$ for $g \equiv \tilde g_A$ and note that, from (1.1), $\mathbb{E}^W D = -\lambda(W-\mu)+R$. This yields

$$\begin{aligned} 0 = \mathbb{E}F(W,W') &= \mathbb{E}\big\{D\big(2g(W)+g(W')-g(W)\big)\big\} \\ &= -2\lambda\mathbb{E}\big\{(W-\mu)g(W)\big\} + 2\mathbb{E}\big\{Rg(W)\big\} + \mathbb{E}\big\{D(g(W')-g(W))\big\}. \end{aligned} \tag{2.13}$$

Note now that, for $D_i := I[D=i]$, $i\in\{-1,+1\}$, we can write

$$D(g(W')-g(W)) = D_{+1}\Delta g(W) + D_{-1}\Delta g(W-1),$$

and further, using exchangeability,

$$\begin{aligned} \mathbb{E}\big\{D_{-1}\Delta g(W-1)\big\} &= \mathbb{E}\big\{I[W'-W=-1]\Delta g(W-1)\big\} \\ &= \mathbb{E}\big\{I[W-W'=1]\Delta g(W')\big\} \\ &= \mathbb{E}\big\{D_{+1}\Delta g(W)\big\}, \end{aligned} \tag{2.14}$$

thus,

$$\mathbb{E}\big\{D(g(W')-g(W))\big\} = 2\mathbb{E}\big\{D_{+1}\Delta g(W)\big\}. \tag{2.15}$$

Together with (2.13) this yields

$$\mathbb{E}\big\{(W-\mu)g(W)\big\} = \frac{\mathbb{E}\big\{D_{+1}\Delta g(W)\big\}}{\lambda} + \frac{\mathbb{E}\big\{Rg(W)\big\}}{\lambda}. \tag{2.16}$$

Note now that, by exchangeability, $\mathbb{E}D_{+1} = \mathbb{E}D_{-1}$ and hence that

$$\begin{aligned} \mathbb{E}D_{+1} &= \tfrac12\mathbb{E}(W'-W)^2 \\ &= \tfrac12\big[\mathbb{E}(W'-\mu)^2 - 2\mathbb{E}\big\{(W'-\mu)(W-\mu)\big\} + \mathbb{E}(W-\mu)^2\big] \\ &= \lambda\sigma^2 + \mathbb{E}\big\{(W-\mu)R\big\} =: \lambda\sigma^2 + a, \end{aligned} \tag{2.17}$$

from (1.1); then use (2.16) to express the expectation in (2.12) as

$$\begin{aligned} &\mathbb{E}\big\{(W-\mu)g(W) - \sigma^2\Delta g(W)\big\} \\ &= \mathbb{E}\big\{(W-\mu)g(W) - (\sigma^2+\lambda^{-1}a)\Delta g(W)\big\} + \lambda^{-1}a\mathbb{E}\Delta g(W) \\ &= \mathbb{E}\big\{\big(D_{+1}\lambda^{-1} - \sigma^2 - \lambda^{-1}a\big)\Delta g(W)\big\} + \lambda^{-1}\mathbb{E}\big\{Rg(W)\big\} + \lambda^{-1}a\mathbb{E}\Delta g(W) \\ &=: B_1 + B_2 + B_3. \end{aligned}$$

Now, recall that $S = \mathbb{E}^W D_{+1}$, and thus, with the estimates

$$|B_1| \leqslant \|\Delta g\|\lambda^{-1}\mathbb{E}\big|S-\mathbb{E}S\big| \leqslant \|\Delta g\|\lambda^{-1}\sqrt{\operatorname{Var} S}, \tag{2.18}$$

$$|B_2| \leqslant \|g\|\lambda^{-1}\mathbb{E}|R| \leqslant \|g\|\lambda^{-1}\sqrt{\operatorname{Var} R}, \tag{2.19}$$

$$|B_3| \leqslant \|\Delta g\|\lambda^{-1}\mathbb{E}\big|(W-\mu)R\big| \leqslant \|\Delta g\|\lambda^{-1}\sigma\sqrt{\operatorname{Var} R}, \tag{2.20}$$

and the bounds (2.7), (2.1) follows.

To prove (2.2), we also use (2.12), but now we take the supremum only over all subsets $A = \{i\}$ for $i \in \mathbb{Z}$. Writing $g \equiv \tilde g_{\{i\}}$ and following the proof as for d_{TV}

above, the bound on (2.19) remains and recalling (2.8), the third term in (2.2) follows. We thus need only refine the bounds on B_1 and B_3. Note that by the Cauchy-Schwartz inequality

$$|B_1| \leqslant \lambda^{-1}\sqrt{\operatorname{Var} S}\sqrt{\mathbb{E}(\Delta g(W))^2}.$$

Using Lemma 2.2, the latter expectation can be bounded by

$$\mathbb{E}(\Delta g(W))^2 = \sum_k (\Delta g(k))^2 \mathbb{P}[W = k] \leqslant q_{\max} \sum_k (\Delta g(k))^2 \leqslant 4\sigma^{-4} q_{\max} \tag{2.21}$$

which implies the first term in (2.2). Using a similar argument on B_3, we obtain

$$|B_3| \leqslant \lambda^{-1}\sigma\sqrt{\operatorname{Var} R}\, q_{\max} \sum_k |\Delta g(k)|,$$

which, together with Lemma 2.2, yields the second term in (2.2). □

Remark 2.1. Theorem 2.1 is a direct analogue of Theorem 1.2 of Rinott and Rotar (1997). However, the first term in (2.1) is slightly different in quality from Theorem 1.2 of Rinott and Rotar (1997), as can be seen by comparing the result of their Theorem 1.3 for the anti-voter model with estimate (3.7). The additional $2/\sigma^2$ in (2.1) and (2.2) occurs because the Possion distribution cannot take negative values, and because the translation must be integer valued. Depending on the problem at hand, this error term can be further reduced or even be omitted; see estimates (2.6) and (2.11).

Remark 2.2. In some of the applications, instead of $S(W) = \mathbb{P}[W' = W + 1|W]$, we will estimate the variance of a random variable $S^* = S^*(X) := \mathbb{P}[W' = W+1|X]$ for some random variable X such that the corresponding σ-algebras satisfy $\sigma(W) \subset \sigma(X)$ and then use the basic fact that $\operatorname{Var} S \leqslant \operatorname{Var} S^*$.

Remark 2.3. As becomes clear from equation (2.16), there is a close connection between the random variable $S = S(W)$ and the so called w-functions as examined for example by Cacoullos et al. (1994) and Cacoullos and Papathanasiou (1997) for the normal and the Poisson distributions. In the case of the standard normal distribution, their problem is as follows: for a given random variable X with $\mathbb{E}X = 0$ and $\operatorname{Var} X = 1$, find a function $w : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{E}\big\{Xf(X)\big\} = \mathbb{E}\big\{w(X)f'(X)\big\} \tag{2.22}$$

holds for a large set of functions f. For the translated Poisson distribution, the corresponding equation is

$$\mathbb{E}\big\{(W-\mu)f(W)\big\} = \mathbb{E}\big\{w(W)\Delta f(W)\big\}, \tag{2.23}$$

and it is indeed satisfied for any W as in Theorem 2.1 if $R = 0$ and if we choose $w(W) = S(W)/\lambda$. Unfortunately, it is often difficult to give an explicit expression for S as a function of W. However, if we allow $w(W)$ in (2.23) to be replaced by a more general random variable, we see from (2.16) that we can use the random variable $S^*(X)/\lambda$ from Remark 2.2 instead. For instance, for the anti-voter model as discussed in the next section, $S^*(X)$ has the nice and explicit representation (3.10).

Using (2.1) with the following corollary one easily obtains a bound for $q_{\max}$.

Corollary 2.3. *For any $\mathbb{Z}$-valued random variable W,*

$$q_{\max} \leqslant d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) + \frac{1}{2.3\sigma}$$

Proof. Just apply Proposition A.2.7 of Barbour et al. (1992). □

Remark 2.4. Estimate (2.2) in combination with Corollary 2.3 is enough to obtain a local limit theorem in the applications of the next section. Although it can be easily calculated in many circumstances, the example of the Poisson-binomial distribution shows that the bound on d_{loc} need not be optimal; estimate (2.2) is of order $\mathrm{O}(n^{-3/4})$ in the special case of the binomial distribution, in contrast to the true order $\mathrm{O}(n^{-1})$. Under additional assumptions on S however, the bound (2.2) can be used to derive the better d_{loc}-bound, given in the following theorem. This bound is used in the examples 3.2 and 3.3 to obtain the correct order $\mathrm{O}(n^{-1})$ of approximation.

Theorem 2.4. *Assume the conditions of Theorem 2.1; assume in addition that S, as a function of W, can be extended on $\mathbb{R}$ such that it is Lipschitz continuous. Then,*

$$\begin{aligned} d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) \leqslant{}& \frac{2L_S\big(\sigma^{-3}\mathbb{E}|W-\mu|^3 \vee (d\sigma^{3/2}+1)\big)}{\lambda\sigma^2} + \frac{2L_S q_{\max}}{\lambda\sigma} \\ &+ \frac{2q_{\max}\sqrt{\operatorname{Var} R}}{\lambda\sigma} + \frac{\sqrt{\operatorname{Var} R}}{\lambda\sigma^2} + \frac{2}{\sigma^2}. \end{aligned} \tag{2.24}$$

where d is the d_{loc}-bound (2.2) *and L_S is the Lipschitz constant of S.*

To prove Theorem 2.4 we need the following Lemma.

Lemma 2.5. *For any μ and σ^2, the bound*

$$\mathrm{TP}(\mu, \sigma^2)\{k\}\,|k-\mu| \leqslant 1$$

holds for all $k \in \mathbb{Z}$.

Proof. Recall from (2.4) that, if $Z \sim \mathrm{TP}(\mu, \sigma^2)$,

$$\mathbb{E}\big\{(Z-\mu)g(Z) - (\sigma^2+\gamma)\Delta g(Z)\big\} = 0 \tag{2.25}$$

for any g for which the expectations exist. With $\pi_k = \mathrm{TP}(\mu, \sigma^2)\{k\}$ and putting $g(\cdot) = I[\cdot = k]$ we obtain from (2.25) the bound

$$\begin{aligned} \pi_k|k-\mu| &\leqslant (\sigma^2+\gamma)|\pi_{k-1} - \pi_k| \\ &\leqslant (\sigma^2+\gamma)\, d_{\mathrm{loc}}\big(\mathrm{TP}(\mu+1, \sigma^2), \mathrm{TP}(\mu, \sigma^2)\big) \\ &= (\sigma^2+\gamma)\, d_{\mathrm{loc}}\big(\mathscr{L}(Y+1), \mathscr{L}(Y)\big). \end{aligned}$$

where $Y \sim \mathrm{Po}(\sigma^2+\gamma)$. The later d_{loc}-distance can easily be bounded using Stein's method for the Poisson distribution, that is, (2.4) in connection with the bound (2.8), which yields $d_{\mathrm{loc}}\bigl(\mathscr{L}(Y+1),\mathscr{L}(Y)\bigr) \leqslant (\sigma^2+\gamma)^{-1}$. □

Proof of Theorem 2.4. Follow the proof of Theorem 2.1 for the d_{loc} metric up to the bounds on the B_i. The bounds on $|\mathrm{B}_2|$ and $|B_3|$ remain. Recalling that S is a function defined on all $\mathbb{R}$, write now B_1 as

$$\begin{aligned} B_1 &= \lambda^{-1}\mathbb{E}\bigl\{(S(W)-\mathbb{E}S(W))\Delta g(W)\bigr\} \\ &= \lambda^{-1}\mathbb{E}\bigl\{(S(W)-S(\mu))\Delta g(W)\bigr\} + \lambda^{-1}\mathbb{E}\bigl\{(S(\mu)-S(W))\bigr\}\mathbb{E}\Delta g(W) \\ &=: B_{1,1}+B_{1,2}. \end{aligned}$$

Exploiting Lipschitz continuity of S and recalling (2.9) we obtain with $q_k = \mathbb{P}[W = k]$

$$|B_{1,2}| \leqslant \lambda^{-1}\sigma L_S \sum_k q_k|\Delta g(k)| \leqslant \frac{2L_S q_{\max}}{\lambda\sigma}$$

which is the second term in (2.24). For $B_{1,1}$ we have

$$\begin{aligned} |B_{1,1}| &\leqslant \lambda^{-1}\sum_k q_k|S(k)-S(\mu)||\Delta g(k)| \\ &\leqslant \lambda^{-1}L_S\sum_k q_k|k-\mu||\Delta g(k)|. \end{aligned} \tag{2.26}$$

We now bound $q_k|k-\mu|$. Assume first that $|k-\mu| > \sigma^{3/2}$; then, by Chebyshev's inequality,

$$q_k \leqslant \mathbb{P}[W \geqslant k] \leqslant \frac{\mathbb{E}|W-\mu|^3}{|k-\mu|^3}$$

and thus

$$q_k|k-\mu| \leqslant \sigma^{-3}\mathbb{E}|W-\mu|^3.$$

On the other hand, if $|k-\mu| \leqslant \sigma^{3/2}$, observe that

$$q_k \leqslant d + \mathrm{TP}(\mu,\sigma^2)\{k\}$$

and hence, using Lemma 2.5,

$$q_k|k-\mu| \leqslant d\sigma^{3/2}+1.$$

Thus, (2.26) can be further bounded to

$$|B_{1,1}| \leqslant \lambda^{-1}L_S\bigl(\sigma^{-3}\mathbb{E}|W-\mu|^3 \vee (d\sigma^{3/2}+1)\bigr)\sum_k|\Delta g(k)|$$

and applying again (2.9), this yields the first term in (2.24). □

The following lemma can be used to estimate the second and third moments of W.

Lemma 2.6. *Under the assumptions of Theorem 2.1 and with* $A = \{w : \mathbb{P}[W = w] > 0\}$ *and* $a := \mathbb{E}\{R(W-\mu)\}$,

$$\lambda^{-1}\Big(\inf_{w\in A} S(w) - a\Big) \leqslant \sigma^2 \leqslant \lambda^{-1}\Big(\sup_{w\in A} S(w) - a\Big),$$
$$\mathbb{E}|W-\mu|^3 \leqslant \lambda^{-1}\big(8q_{\max} + 1 + \sigma + \mathbb{E}\{|R|(W-\mu)^2\}\big).$$

Proof. The estimates for the variance are immediate from equality (2.17) and the bounds

$$\inf_{w\in A} S(w) \leqslant \mathbb{E}S(W) \leqslant \sup_{w\in A} S(w).$$

Note now that from equation (2.16),

$$\mathbb{E}\big\{(W-\mu)g(W)\big\} = \lambda^{-1}\mathbb{E}\big\{S(W)\Delta g(W)\big\} + \lambda^{-1}\mathbb{E}\big\{Rg(W)\big\}$$

for all functions g, for which the expectations exist. With $K_\mu(w) = I[w > \mu] - I[w \leqslant \mu]$ and $g(w) = K_\mu(w)(w-\mu)^2$ we thus obtain

$$\begin{aligned}\mathbb{E}|W-\mu|^3 &= \lambda^{-1}\mathbb{E}\big\{S(W)\big[(W-\mu)^2 + 2(W-\mu) + 1\big]\Delta K_\mu(W)\big\}\\ &\quad + \lambda^{-1}\mathbb{E}\big\{S(W)(2(W-\mu)+1)K_\mu(W)\big\}\\ &\quad + \lambda^{-1}\mathbb{E}\big\{R(W-\mu)^2K_\mu(W)\big\} =: B_1' + B_2' + B_3'\end{aligned}$$

Note now, that $|K(w)| = 1$ and

$$\Delta K_\mu(w) = \begin{cases} 2 & \text{if } w = \lfloor\mu\rfloor,\\ 0 & \text{else,}\end{cases}$$

and thus, as $|\lfloor\mu\rfloor - \mu| \leqslant 1$ and $|S(w)| \leqslant 1$,

$$|B_1'| \leqslant 8\lambda^{-1}q_{\max},$$
$$|B_2'| \leqslant \lambda^{-1} + \lambda^{-1}\sigma.$$

The bound on B_3' is immediate. □

3. Applications

In this section we illustrate our results using some examples in in which $W = \sum_{i=1}^n J_i$ for a sequence $J = (J_1, J_2, \ldots, J_n)$ of random indicators. Barbour and Xia (1999) and Röllin (2005) considered cases where the J_i have a local dependence structure; in contrast, the examples in this paper (with the exception of the first, standard example) exhibit global dependence.

For latter use we recall the following easy to prove fact.

Lemma 3.1. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a Lipschitz continuous function with Lipschitz constant* L_f. *Then, for any random variable* X,

$$\operatorname{Var} f(X) \leqslant L_f^2 \operatorname{Var} X.$$

3.1. Poisson-Binomial distribution.

Theorem 3.2. *Let $J = (J_1, \dots, J_n)$ be a sequence of independent random indicators with $\mathbb{E}J_i = p_i$. Then, Theorem 2.1 can be applied with $R = 0$ and $\lambda = 1/n$; we have*

$$S^*(J) := \mathbb{P}[W' = W + 1|J] = \frac{1}{n}\sum_{i=1}^{n}(1 - J_i)p_i, \tag{3.1}$$

$$\operatorname{Var} S(W) \leqslant \operatorname{Var} S^*(J) = n^{-2}\sum_{i=1}^{n} p_i^3(1 - p_i), \tag{3.2}$$

$$q_{\max} \leqslant 0.47\sigma^{-1}, \tag{3.3}$$

where (3.3) *holds if $\sigma^2 = \sum_{i=1}^{n} p_i(1 - p_i) \geqslant 4$. Thus, if $\sigma^2 \asymp n$,*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) = \mathrm{O}(n^{-1/2})$$
$$d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) = \mathrm{O}(n^{-3/4}).$$

Remark 3.1. As already mentioned in Remark 2.4, the order of the above d_{loc}-bound is not optimal. However, Röllin (2006) obtains the correct order $\mathrm{O}(n^{-1})$, using a different variant of Stein's method where an explicit smoothing argument is involved.

Remark 3.2. Corollary 2.1 of Čekanavičius and Vaĭtkus (2001) seems to be better in constant than (2.1) in the above example. For instance, for the Binomial distribution, we have

$$d_{\mathrm{TV}}\big(\mathrm{Bi}(n,p), \mathrm{TP}(\mu, \sigma^2)\big) \leqslant C\sqrt{\frac{p}{n(1-p)}} + \frac{2}{np(1-p)},$$

where Čekanavičius and Vaĭtkus (2001) obtain $C = 0.93$ and (2.1) yields $C = 1$.

Proof. We use the standard argument from Stein (1986). Let K be uniformly distributed on $\{1, \dots, n\}$ and let J^* be an independent copy of J. Then it is easy to see that $W' := W - J_K + J_K^*$ will satisfy (1.1) with $R = 0$ and $\lambda = 1/n$. Further

$$\begin{aligned} S^*(J) &= \mathbb{E}^J I[W' - W = 1] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}^J I[J_i = 0, J_i^* = 1] \\ &= \frac{1}{n}\sum_{i=1}^{n}(1 - J_i)\mathbb{E}^J J_i^* = \frac{1}{n}\sum_{i=1}^{n}(1 - J_i)p_i \end{aligned}$$

Thus, $\operatorname{Var} S^* = n^{-2}\sum_{i=1}^{n} p_i^3(1 - p_i)$, which, following Remark 2.2, proves (3.2). From (Čekanavičius and Vaĭtkus, 2001, estimate (2.22)) we obtain

$$q_{\max} \leqslant 0.45\sigma^{-1}(16/15)^{1/2},$$

if $\sigma^2 \geqslant 4$ which proves (3.3). □

3.2. Hypergeometric distribution. Assume that we have N urns and m balls, and that we distribute the balls uniformly into the N urns, in such a way that there is at most one ball per urn. Clearly, the number of balls W in the first n urns has the hypergeometric distribution $\mathrm{Hyp}(m,n,N)$, for which

$$\sigma^2 = \operatorname{Var} W = \frac{nm(N-n)(N-m)}{(N-1)N^2}.$$

Theorem 3.3. *If W has the hypergeometric distribution, Theorem 2.1 can be applied with $R=0$ and $\lambda = \frac{N}{m(N-m+1)}$; we have*

$$\begin{aligned} S(W) &= \frac{mn-(m+n)W+W^2}{m(N-m+1)}, \\ \operatorname{Var} S(W) &\leqslant \frac{nm(m+n)^2(N-n)(N-m)}{m^2(N-m+1)^2(N-1)N^2}; \end{aligned} \tag{3.4}$$

thus, if $N=N(n)\asymp n$ and $m=m(n)\asymp n$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\mathrm{TP}(\mu,\sigma^2)\big) = \mathrm{O}(n^{-1/2}),$$
$$d_{\mathrm{loc}}\big(\mathscr{L}(W),\mathrm{TP}(\mu,\sigma^2)\big) = \mathrm{O}(n^{-1}).$$

Proof. Consider the following construction. Pick uniformly an urn with a ball, and put this ball into any empty urn (including the urn, from which the ball was picked). Denote now by W' the number of balls in the first n urns. Exchangeability of (W,W') is easy to see and condition (1.2) is clearly satisfied. Now, $W'-W=1$ is the event that a ball is picked from one of the urns $n+1,\dots,N$ and put into one of the empty urns $1,\dots,n$, thus

$$S(W) = \mathbb{P}[W'=W+1|W] = \frac{m-W}{m}\times\frac{n-W}{N-m+1}$$

and conversely

$$\mathbb{P}[W'=W-1|W] = \frac{W}{m}\times\frac{N-n-m+W}{N-m+1}$$

thus

$$\begin{aligned} &\mathbb{E}^W(W'-W) \\ &\quad = \mathbb{E}^W I[W'-W=1] - \mathbb{E}^W I[W'-W=-1] = \frac{mn-NW}{m(N-m+1)}, \end{aligned}$$

and (1.1) is satisfied with $R=0$ and $\lambda = \frac{N}{m(N-m+1)}$.

Note now that S, as a function of W, is Lipschitz continuous with constant $L_S = \frac{m+n}{m(N-m+1)}$; thus applying Lemma 3.1 we have

$$\operatorname{Var} S \leqslant \frac{(m+n)\sigma^2}{m^2(N-m+1)^2}. \qquad \square$$

This is enough to prove the d_{TV}-order and, together with Corollary 2.3, the order $\mathrm{O}(n^{-3/4})$ for the d_{loc}-metric. Now, noting that Lemma 2.6 yields $\mathbb{E}|W-\mu|^3 = \mathrm{O}(n^{3/2})$, we obtain from Theorem 2.4 the desired order $\mathrm{O}(n^{-1})$ for the d_{loc}-metric.

3.3. A parity problem. Let $J_1, \dots, J_n$ be a sequence of independent Be(1/2)-distributed random indicators. Define

$$J_{n+1} := \begin{cases} 1 & \text{if } \sum_{i=1}^n J_i \text{ is odd,} \\ 0 & \text{else,} \end{cases}$$

and $V := \sum_{i=1}^{n+1} J_i$, so V is simply obtained by 'rounding' a $\mathrm{Bi}(n, 1/2)$-distributed random variable to the next even integer. An approximation of V by a translated Poisson distribution will clearly not succeed; however, we may try with $W := \frac{1}{2}V$.

Regard now the following exchangeable pair coupling. Pick two random indices K, L uniformly on $\{1, \dots, n+1\}$ so that almost surely $K \neq L$, and define

$$V' = V + 2 - 2J_K - 2J_L; \tag{3.5}$$

that is, take two summands of V at random, and replace each of them by its complement.

Lemma 3.4. *The pair (V, V') defined as above is a exchangeable and $(W, W') := (\frac{1}{2}V, \frac{1}{2}V')$ satisfies (1.1) and (1.2) with $\lambda = 2/(n+1)$.*

Proof. It is enough to regard the situation on $M = \{0,1\}^n$ because the values $J_1, \dots, J_n$ uniquely determine the random variable J_{n+1}. Note first that construction (3.5) gives rise to a discrete time Markov chain on M, with jumps from $j \in M$ to $j' \in M$, if j' differs from j in exactly one or two coordinates (j' differing in exactly one coordinate corresponds to K or L being equal to $n+1$). Now, as the jump from j to j' happens with the same probability as from j' to j and all the states are connected, it is easy to see that the such defined Markov chain is irreducible and reversible and that the equilibrium distribution assigns equal probability to any $j \in M$, which corresponds to n independent Be(1/2) random variables. Thus, exchangeability is proved.

Note now that

$$\begin{aligned} \mathbb{E}^J(V' - V) &= 2 - \frac{2}{n(n+1)} \sum_{k=1}^{n+1} \sum_{\substack{l=1 \\ l \neq k}}^{n+1} (J_k + J_l) \\ &= 2 - \frac{2}{n(n+1)} 2nV = 2 - \frac{4V}{n+1}, \end{aligned}$$

thus, we can take $\lambda = 2/(n+1)$. ☐

Theorem 3.5. *For W defined as above, Theorem 2.1 can be applied with $R = 0$ and $\lambda = 2/(n+1)$; if $n \geqslant 2$, we have $\sigma^2 = (n+1)/16$ and*

$$S(W) = \frac{n(n+1) - (4n-2)W + 4W^2}{n(n+1)}, \quad \operatorname{Var} S(W) \leqslant \frac{(4n-2)^2(n+1)}{16n^2(n+1)^2};$$

thus, as $n \to \infty$,

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) &= \mathrm{O}(n^{-1/2}), \\ d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) &= \mathrm{O}(n^{-1}). \end{aligned}$$

Proof. Note first that if $n \geqslant 2$ the J_i are uncorrelated and thus

$$\sigma^2 = \operatorname{Var}(V)/4 = (n+1)/16.$$

Now,

$$\begin{aligned}\mathbb{E}^J I[W' - W = 1] &= \frac{1}{n(n+1)} \sum_{k=1}^{n+1} \sum_{\substack{l=1 \\ l \neq k}}^{n+1} (1 - J_k)(1 - J_l) \\ &= \frac{n(n+1) - (4n-2)W + 4W^2}{n(n+1)} =: S(W).\end{aligned}$$

Observe that S, as a function of W, is Lipschitz continuous with $L_S = \frac{4n-2}{n(n+1)}$; thus, applying Lemma 3.1,

$$\operatorname{Var} S(W) \leqslant \frac{(4n-2)^2\sigma^2}{n^2(n+1)^2}.$$

This is enough to prove the d_{TV}-order and, together with Corollary 2.3, the order $O(n^{-3/4})$ for d_{loc}. Now, noting that Lemma 2.6 yields $\mathbb{E}|W - \mu| = \mathrm{O}(n^{3/2})$, we obtain from Theorem 2.4 the desired order $\mathrm{O}(n^{-1})$ for the d_{loc}-metric.

□

3.4. Anti-voter model on finite graphs. We closely follow the setup of Rinott and Rotar (1997); see also references therein and Huber and Reinert (2004). Let G be a n-vertex r-regular graph, which is neither bipartite nor an n-cycle. At each vertex i we assume that there is a 'voter' attached, having an opinion $J_i^{(t)}$ which can take the values 0 or 1 in every time point $t \in \mathbb{N}$. Define a Markov chain by the following transition rule. Choose uniformly a random vertex, say i; then, out of the neighbourhood $\mathcal{N}_i$ of i, choose uniformly a random vertex, say j, and let $J_i^{(t+1)}$ be the opposite of $J_j^{(t)}$ and leave the other voters untouched. Assume now that the Markov chain is in its equilibrium and put $W = \sum_{i=1}^n J_i := \sum_{i=1}^n J_i^{(0)}$.

Theorem 3.6. *For the anti-voter model as described above, Theorem 2.1 can be applied with $R = 0$ and $\lambda = 2/n$; we have*

$$S^*(J) := \mathbb{E}^J I[W' - W = 1] = \frac{3rn - 4rW + Q}{4rn}, \tag{3.6}$$

$$\operatorname{Var} S(W) \leqslant \operatorname{Var} S^*(J) \leqslant \frac{16r^2\sigma^2 + \operatorname{Var} Q}{16r^2n^2}, \tag{3.7}$$

where

$$Q = \sum_{i=1}^n \sum_{j \in \mathcal{N}_i} (2J_i - 1)(2J_j - 1);$$

hence, as $n \to \infty$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) = \mathrm{O}\left(\frac{\sqrt{\operatorname{Var} Q}}{r\sigma^2} + \frac{1}{\sigma}\right),$$
$$d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) = \mathrm{O}\left(\frac{(\operatorname{Var} Q)^{3/4}}{r^{3/2}\sigma^3} + \frac{1}{\sigma^{3/2}}\right).$$

Remark 3.3. Note that the bounds for d_{TV} in Theorem 3.6 is very similar to the bound for the weaker Kolmogorov metric d_{K} given in Theorem 1.3 of Rinott and Rotar (1997); they obtain

$$d_{\mathrm{K}}\big(\mathscr{L}(W_c), \mathcal{N}(0,1)\big) = \mathrm{O}\left(\frac{\sqrt{\operatorname{Var} Q}}{r\sigma^2} + \frac{n}{\sigma^3}\right), \tag{3.8}$$

where $W_c = (W - \mu)/\sigma$.

Example 3.7. Consider the sequence K_n of complete graphs of size n. Rinott and Rotar (1997) show that $\sigma^2 \asymp n$ and $\operatorname{Var} Q = \mathrm{O}(n^3)$. Thus, from Theorem 2.1, the d_{TV}-distance is of the order $\mathrm{O}(n^{-1/2})$ and the d_{loc}-distance of order $\mathrm{O}(n^{-3/4})$ which proves the LLT. Now,

$$S^*(J) = \frac{n(n-1) - (2n-1)W + W^2}{n(n-1)} = S(W), \tag{3.9}$$

and we can thus take $L_S = \frac{2}{n-1}$. From Lemma 2.6 we obtain $\mathbb{E}|W - \mu|^3 = \mathrm{O}(n^{3/2})$ and therefore Theorem 2.4 yields the order $\mathrm{O}(n^{-1})$ for d_{loc}. Note that the estimates on L_S is obtained only because of the explicit representation (3.9); they are difficult to obtain in general. For further examples of graphs see Rinott and Rotar (1997).

Proof of Theorem 3.6. Define $W' := \sum_{i=1}^n J_i^{(1)}$, and note that (W, W') is an exchangeable pair, satisfying (1.1) and (1.2) with the choices $\lambda = 2/n$ and $R = 0$ (for more details see Rinott and Rotar (1997)). Now, let K be the random index of the vertex that was resampled in the transition from W to W'. As $W' = W - J_K + J_K^{(1)}$,

$$\begin{aligned} S^*(J) = \mathbb{E}^J I[W' - W = 1] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}^J\big\{I[J_i = 0, J_i^{(1)} = 1] \bigm| K = i\big\} \\ &= \frac{1}{n}\sum_{i=1}^n (1 - J_i)\mathbb{E}^J\big\{J_i^{(1)} \bigm| K = i\big\} = \frac{1}{n}\sum_{i=1}^n (1 - J_i)\Big(1 - \frac{1}{r}\sum_{j \in \mathcal{N}_i} J_j\Big) \end{aligned} \tag{3.10}$$

With $X_i = 2J_i - 1$ and $\tilde{W} = \sum_{i=1}^n X_i$, (3.10) becomes

$$\begin{aligned} S^*(J) &= \frac{1}{4rn}\sum_{i=1}^n (1 - X_i)\Big(r - \sum_{j \in \mathcal{N}_i} X_j\Big) \\ &= \frac{1}{4rn}\Big(rn - \sum_{i=1}^n \sum_{j \in \mathcal{N}_i} X_j - r\sum_{i=1}^n X_i + Q\Big) \\ &= \frac{rn - 2r\tilde{W} + Q}{4rn}. \end{aligned}$$

The variance of S^* is thus

$$\operatorname{Var} S^*(J) = \frac{\operatorname{Var}(2r\tilde{W}) + \operatorname{Var} Q - 4r\operatorname{Cov}(\tilde{W}, Q)}{16r^2n^2} = \frac{16r^2\sigma^2 + \operatorname{Var} Q}{16r^2n^2}, \tag{3.11}$$

because $\mathbb{E}\{X_iX_jX_k\} = 0$ for any choice of i, j and k, due to the symmetry of the anti-voter model, and hence $\mathbb{E}\{\tilde{W}Q\} = 0$. $\square$

Acknowledgements

I thank Andrew Barbour for many helpful suggestions and comments.

References

A. D. Barbour and T. Lindvall (to appear). Translated Poisson approximation for Markov chains. *J. Theor. Probab.*.

A. D. Barbour, L. Holst, and S. Janson (1992). *Poisson approximation.* The Clarendon Press Oxford University Press, New York.

A. D. Barbour and A. Xia (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3**, 131–150.

T. Cacoullos and V. Papathanasiou (1997). Characterizations of distributions by generalizations of variance bounds and simple proofs of the CLT. *J. Statist. Plann. Inference* **63**, 157–171.

T. Cacoullos, V. Papathanasiou, and S. A. Utev (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

V. Čekanavičius and P. Vaĭtkus (2001). Centered Poisson approximation by the Stein method. *Lithuanian Math. J.* **41**, 319–329.

S. Chatterjee, P. Diaconis, and E. Meckes (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2**, 64–106.

M. Huber and G. Reinert (2004). The stationary distribution in the antivoter model: exact sampling and approximations. In *Stein's method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, 79–93.

Y. Rinott and V. Rotar (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7**, 1080–1105.

A. Röllin (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.

A. Röllin (2006, submitted manuscript). Symmetric binomial approximation for sums of locally dependent random variables.

C. Stein (1986). *Approximate computation of expectations*, IMS Lecture Notes. Institute of Mathematical Statistics, Hayward.